Dynamic and Stochastic Rational Behavior*

Nail Kashaev ⓡ Victor H. Aguiar ⓡ Martin Plávala ⓡ Charles Gauthier†

April, 2023

Abstract We analyze choice behavior using Dynamic Random Utility Model (DRUM). Under DRUM, each consumer or decision-maker draws a utility function from a stochastic utility process in each period and maximizes it subject to a menu. DRUM allows for unrestricted time correlation and cross-section heterogeneity in preferences. We fully characterize DRUM when panel data on choices and menus are available. Our results cover consumer demand with a continuum of choices and finite discrete choice setups. DRUM is linked to a finite mixture of deterministic behaviors that can be represented as the Kronecker product of static rationalizable behaviors. We exploit a generalization of the Weyl-Minkowski theorem that uses this link and enables conversion of the characterizations of the static Random Utility Model (RUM) of McFadden-Richter (1990) to its dynamic form. DRUM is more flexible than Afriat's (1967) framework and more informative than RUM. In an application, we find that static utility maximization fails to explain population behavior, but DRUM can explain it.

JEL classification numbers: C10, C33, D11, D12, D15.

Keywords: dynamic random utility, revealed preference.

*This paper subsumes "Nonparametric Analysis of Dynamic Random Utility Models." The "ⓡ" symbol indicates that the authors' names are in certified random order, as described by Ray and Robson (2018). We thank Roy Allen, Chris Chambers, Pierre-André Chiappori, Mark Dean, Adam Dominiak, Laura Doval, David Freeman, Matt Kovach, Elliot Lipnowski, Paola Manzini, Krishna Pendakur, Jörg Stoye, Tomasz Strzalecki, and Levent Ülkü for useful discussions and encouragement. Plávala acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project numbers 447948357 and 440958198), the Sino-German Center for Research Promotion (Project M-0294), the ERC (Consolidator Grant 683107/TempoQ), the German Ministry of Education and Research (Project QuKuK, BMBF Grant No. 16KIS1618K), and the Alexander von Humboldt Foundation.

†Kashaev: Department of Economics, University of Western Ontario; nkashaev@uwo.ca. Aguiar: Department of Economics, University of Western Ontario; vaguiar@uwo.ca. Plávala: Naturwissenschaftlich-Technische Fakultät, Universität Siegen; martin.plavala@uni-siegen.de. Gauthier: ECARES, Université Libre de Bruxelles; charles.gauthier@ulb.be.

1. Introduction

A fundamental question in economics is whether decision makers exhibit rational choice behavior. Traditional definitions of rationality are effectively equivalent to maximizing a utility function that is fixed in time. Here, we study a notion of rationality in choice behavior that is stochastic and dynamic—Dynamic Random Utility Model (DRUM). Under DRUM, each consumer or decision maker (DM) at each period maximizes the realized utility from a stochastic utility process subject to a menu or budget. We provide a revealed preference characterization of DRUM when the longitudinal distribution of choices or demands is observed for a finite collection of menus or budgets in a finite time window. This characterization does not make any parametric restriction on (i) the form of utility functions, (ii) the correlation of utilities in time, and (iii) the heterogeneity of utility in the cross-section.

There are two main frameworks to analyze consumer behavior: Samuelson-Afriat (1938, 1967)'s framework of static utility maximization and McFadden-Richter (1990)'s framework of random utility maximization (RUM). DRUM addresses several empirical limitations of these models. In particular, Samuelson-Afriat's framework is under scrutiny due to experimental and field evidence against it.[1] There is evidence that failures of Samuelson-Afriat's framework are driven by the stringent assumption of the stability of preferences over time. For example, utility functions may change over time because of variability in time of the neural computation of value (Kurtz-David, Persitz, Webb and Levy, 2019), structural breaks (Cherchye, Demuynck, De Rock and Vermeulen, 2017), or evolving risk aversion (Guiso, Sapienza and Zingales, 2018, Akesaka, Eibich, Hanaoka and Shigeoka, 2021). DRUM allows preferences to change freely in time. In contrast to Samuelson-Afriat's framework, RUM has found reasonable success

[1] For examples in household consumption see Echenique, Lee and Shum (2011), Dean and Martin (2016) and in choices over portfolios over risk or uncertainty see Choi, Fisman, Gale and Kariv (2007), Choi, Kariv, Müller and Silverman (2014), Ahn, Choi, Gale and Kariv (2014). The rationality violations were originally thought to be small (Echenique *et al.*, 2011, Choi *et al.*, 2007), but newer experimental data sets show these violations can be severe (Brocas, Carrillo, Combs and Kodaverdian, 2019, Aguiar and Serrano, 2021, Halevy and Mayraz, 2022).

explaining repeated cross-sections of household choices (Kawaguchi, 2017, Kitamura and Stoye, 2018). However, RUM cannot take advantage of the longitudinal variation in choices available in many datasets, and it may have limited empirical bite (Im and Rehbeck, 2021). By considering a richer primitive, we simultaneously relax the assumption of stable preferences over time implicit in Samuelson-Afriat's framework while providing a more informative test of stochastic utility maximization than in McFadden-Richter's framework.

Our first result is a mixture characterization of DRUM, which is analogous to the RUM characterization in McFadden-Richter's work. We exploit that DRUM is associated with a finite mixture of preference profiles in time. We obtain results analogous to Kitamura and Stoye (2018) (henceforth KS), McFadden and Richter (1990), and Kawaguchi (2017) with a dynamic version of the Axiom of Stochastic Revealed Preferences. This finite mixture characterization lends itself to statistical testing using results in KS. Also, this characterization can be used for nonparametric counterfactual analysis. In a Monte Carlo study, we show that the statistical test of KS applied to our characterization of DRUM performs well in finite samples.

We show that the mixture representation of DRUM can be obtained using a Kronecker product of the mixture representation of RUM in each period.[2] This observation is vital to obtain: (i) computational gains for testing because of the modularity of the mixture representation; and (ii) a novel characterization of DRUM using a recursive version of the Block and Marschak (1960) inequalities (BM inequalities). In a static setting, KS were the first to observe that the empirical content of RUM can be expressed as cone restrictions on observed data. In particular, the Weyl-Minkowski theorem posits that a cone can be described equivalently by a convex combination of its vertices ($\mathcal{V}$-representation) or by its faces ($\mathcal{H}$-representation). KS note that the $\mathcal{H}$-representation corresponds to what decision theorists would call an axiomatic

[2] Informally, the mixture representation of RUM can be represented as a matrix whose columns are deterministic rational demand types. The analogous matrix for DRUM is the Kronecker product of those RUM matrices.

characterization of RUM.[3] We exploit recent mathematical advancements in the analysis of the Kronecker products of cones (Aubrun, Lami, Palazuelos and Plávala, 2021, Aubrun, Müller-Hermes and Plávala, 2022) to provide an axiomatic characterization of DRUM using the axiomatic characterization of RUM. We bring this new mathematical tool to economics and show how it can be used in the DRUM setup and in structurally similar models.[4]

The generalized Weyl-Minkowski theorem enables us to provide a full characterization of DRUM via dynamic BM inequalities, covering as a special case the finite abstract setup of Li (2021) and Chambers, Masatlioglu and Turansick (2021) with full menu variation. Our characterization works for cases of limited observability of menus and in the presence of a primitive order that the support of the utility process respects. We also provide a novel behavioral condition necessary for consistency of the longitudinal distribution of demand with DRUM (D-monotonicity). It is also sufficient in simple setups: (i) for any finite number of goods and 2 budgets per period, and (ii) for 2 goods and any finite number of budgets per period. D-monotonicity is computationally simple to check and provides a deeper understanding of the empirical content of DRUM. It restricts the joint probability of choices in time beyond the RUM restrictions on marginal distributions in each period. D-monotonicity can be thought of as a dynamic version of the Weak Axiom of Stochastic Revealed Preference (Bandyopadhyay, Dasgupta and Pattanaik, 1999, Hoderlein and Stoye, 2014) and a stochastic version of the Weak Axiom of Revealed Preference (in time series) by Samuelson (1938).

We synthesize the two main paradigms of nonparametric demand analysis, Samuelson-Afriat's and McFadden-Richter's frameworks. Samuelson-Afriat's framework requires observing a *time-series* of choices and budgets for a given consumer and assumes that the consumer maximizes the same utility function each period. When preferences are allowed to vary in time, there are no empirical implications with only a time-series of choices. However, using a *panel*

[3] Note that the BM inequalities are the $\mathcal{H}$-representation of RUM in the finite abstract setup.

[4] For example, models of bounded rationality, such as a model of random consideration of Cattaneo, Ma, Masatlioglu and Suleymanov (2020) and its extension with heterogeneous preferences in Kashaev and Aguiar (2022a), can be extended to a dynamic setup in the spirit of DRUM.

of choices, DRUM bounds the share of consumers or DMs whose choices contain a revealed preference violation in the Afriat's sense. RUM instead requires observing a *cross-section* of choices and budgets from a population of consumers. The panel structure is ignored as there is no time dimension. Hence, this approach misses the potential temporal correlation of utilities. As a result, there are panels of choices over menus that, when marginalized, are consistent with RUM, but not with DRUM. In other words, ignoring the time dimension of choices may lead to false positives when testing DRUM. Importantly, our setup keeps the fundamental assumption in McFadden-Richter's framework—the distribution of utilities does not depend on the sequence of budgets or menus that the consumer faces in time.[5]

Our synthesis is advantageous because it (i) provides more informative bounds on counterfactual choice due to the richer variation in the panel of choices; (ii) provides a theoretical justification for marginalizing choices and using the RUM framework; and (iii) clarifies the role of constant preferences across time in Samuelson-Afriat's framework. Fortunately, our primitive with a longitudinal level of variation is readily available in many consumption surveys, household scanner datasets, and experimental datasets as documented in Aguiar and Kashaev (2021).[6] In our application, we find support for DRUM in an experimental data set of panel choices collected by Aguiar, Boccardi, Kashaev and Kim (2023) (henceforth ABBK). Specifically, we find that static utility maximization fails to explain the behavior of the sample of DMs, yet DRUM can explain the data. Monte Carlo experiments mimicking the application setup show that the power of our DRUM test is high in finite sample.

The DRUM framework is rich and extends well beyond the Samuelson-Afriat and McFadden-Richter worlds. We cover as special cases: (i) consumption models of errors in the evaluation of utility (Kurtz-David *et al.*, 2019); (ii) dynamic random expected utility (defined in Frick, Iijima and Strzalecki, 2019); (iii) static utility maximization in a population (without measurement

[5]This assumption can be relaxed in the same spirit as Deb, Kitamura, Quah and Stoye (2021).

[6]In practice, panels of choices are often pooled in the time dimension to create a cross-section with sufficient budget variations (Deb *et al.*, 2021, Kitamura and Stoye, 2018). In this case, we show that this approach could lead to false rejections of DRUM due to ignoring the time labels of budgets.

error) (Aguiar and Kashaev, 2021); (iv) dynamic utility maximization in a population (Browning, 1989, Gauthier, 2018, Aguiar and Kashaev, 2021); (v) changing utility or multiple-selves models (Cherchye *et al.*, 2017); and changing-taste modelled with a constant utility in time with an additive shock (Adams, Blundell, Browning and Crawford, 2015).

Related Literature. DRUM was first introduced in Strzalecki (2021) for abstract discrete choice domains. Frick *et al.* (2019) offered an axiomatic characterization with decision trees and expected utility restrictions on stochastic utility processes.[7] A BM-like DRUM characterization remained open until our work. In finite abstract discrete choice spaces, two partial characterizations exist when the primitive is the joint distribution of choices across time and total menu variation.

Li (2021) provides an axiomatic DRUM characterization (analogous to BM inequalities for RUM, Block and Marschak, 1960) for any finite number of periods, full menu variation, but no more than 3 alternatives. Chambers *et al.* (2021) considers correlated choice, a joint distribution of choice on a pair of menus faced by two different DMs or a group. This model is mathematically equivalent to DRUM in the abstract domain, characterizing DRUM for a special case of 2 periods, an abstract and finite choice set, but with a uniqueness property for one DM's choice (i.e., one period has uniquely identified RUM). Our work subsumes and generalizes these results. Moreover, our results go beyond both Li (2021) and Chambers *et al.* (2021), and our general setup also includes classical consumer choice domains with primitive orders, binary menus, and general limited menu variation. Our DRUM respects this order and restricts utilities to be monotone, addressing limited observability of menus and menu paths.

Our work also contributes to the random exponential discounting literature, as in Browning (1989), generalizing Deb *et al.* (2021)'s demand setup to a dynamic context. Apesteguia, Ballester and Gutierrez-Daza (2022) introduces a heterogeneous model with exponential discounting and time separability, differing in choice domains and being semiparametric, while

[7]Applying Kashaev and Aguiar (2022b) techniques, we provide a KS-type characterization for special DRUM cases with expected utility restrictions.

our setup is nonparametric. Lu and Saito (2018) examines exponential discounting with random discount factors and stochastic choices over consumption streams in the first period.

Aguiar and Kashaev (2021) studies panel setups with first-order-conditions approaches for some dynamic preferences, allowing for measurement error. However, their setup doesn't accommodate changing utility beyond discount factors or marginal utility of income. Im and Rehbeck (2021) investigates McFadden-Richter's framework and panel structure limitations, suggesting individual static rationality checks, like Samuelson-Afriat's framework. We generalize Afriat's framework, allowing utility changes over time while exploiting panel structure for more empirical implications.

DRUM allows for consumption to be correlated in time, similar to rational addiction in Becker and Murphy (1988) (referred to as Habits as Durables–HAD). However, consumption correlation in DRUM is solely due to preference correlation over time. Demuynck and Verriest (2013) provides an Afriat's theorem-like characterization of the short-memory habits model of Becker and Murphy (1988) and applies it to a panel of Spanish households' choices. However, the HAD pass-rate is slightly above 50%, subject to the same critique as the traditional utility maximization problem. This revealed-preference test of HAD imposes no parametric restrictions on utilities besides monotonicity and concavity, making it more general than parametric structural work on HAD. Theoretically, the relationship between DRUM and HAD remains an open question since a RUM-like characterization of a stochastic generalization of HAD is unavailable.

Dynamic Discrete Choice (DDC) models, surveyed in Aguirregabiria and Mira (2010), are the most popular approach in discrete choice with attribute variation. Frick *et al.* (2019) thoroughly explores the relationship between DRUM and DDC in their domain. Much of this analysis carries over to our domain. Specifically, in our setup with exogenously given menu paths, DDC is nested by DRUM. However, our focus differs from most work in DDC. We emphasize nonparametric utilities, unrestricted heterogeneity, comparative statics, and counterfactual predictions, rather than identification and estimation in parametric settings.

There is no axiomatization of DDC except for the special case of independent and identically distributed logit shocks in Fudenberg and Strzalecki (2015), so we cannot fully compare DDC with DRUM.

Outline. The paper is organized as follows. Section 2 introduces the setup. Section 3 provides a McFadden-Richter and KS-type characterization of DRUM. Section 4 provides a behavioral characterization of DRUM via linear inequality constraints. Section 5 synthesizes Afriat's and McFadden-Richter's setups. Section 6 provides results about dynamic counterfactual analysis. Section 7 provides an application to experimental data. Section 8 provides Monte Carlo experiments showcasing the finite sample properties of the econometric test of DRUM. Section 9 concludes. All proofs can be found in Appendix 10.

2. Setup

We consider a time window $\mathcal{T} = \{1, \cdots, T\}$ with a finite terminal period $T \geq 1$. Let X^t be a nonempty finite choice set. We endow $2^{X^t} \setminus \{\emptyset\}$ with some acyclic partial order $>^t$. When $>^t$ is restricted to singletons it induces an acyclic partial order on X^t. We will abuse notation and write $x >^t y$ instead of $\{x\} >^t \{y\}$ in this case. In each $t \in \mathcal{T}$, there are $J^t < \infty$ distinct menus

$$B^t_j \in 2^{X^t} \setminus \{\emptyset\}, \quad j \in \mathcal{J}^t = \{1, \ldots, J^t\}.$$

Since X^t is a finite set, we denote the i-th element of menu $j \in \mathcal{J}^t$ as $x^t_{i|j}$. That is, $B^t_j = \{x^t_{i|j}\}_{i \in \mathcal{I}^t_j}$, where $\mathcal{I}^t_j = \{1, 2, \ldots, I^t_j\}$ and I^t_j is the number of elements in menu j.

Define a menu path as an ordered collection of indexes $\mathbf{j} = (j_t)_{t \in \mathcal{T}}$, $j_t \in \mathcal{J}^t$. Menu paths encode menus that were faced by agents in different time periods. Let $\mathbf{J}$ be the set of all *observed* menu paths. Given $\mathbf{j} \in \mathbf{J}$, a *choice path* is an array of alternatives $x_{\mathbf{i}|\mathbf{j}} = \left(x^t_{i_t|j_t}\right)_{t \in \mathcal{T}}$ for some collection of indexes $\mathbf{i} = (i_t)_{t \in \mathcal{T}}$ such that $i_t \in \mathcal{I}^t_{j_t}$ for all t. Similar to a menu path, a choice path encodes choices of a DM in a given sequence of menus she faced. The set of all possible choice path index sets $\mathbf{i}$, given a menu path $\mathbf{j}$, is denoted by $\mathbf{I_j}$.

Note that every $\mathbf{j} \in \mathbf{J}$ encodes the Cartesian product of menus $\times_{t\in\mathcal{T}} B^t_{j_t} \subseteq \times_{t\in\mathcal{T}} X^t$. Then, for every $\mathbf{j}$ let $\rho_{\mathbf{j}}$ be a probability measure on $\times_{t\in\mathcal{T}} B^t_{j_t}$. That is, $\rho_{\mathbf{j}}\left(x_{\mathbf{i}|\mathbf{j}}\right) \geq 0$ for all $\mathbf{i} \in \mathbf{I_j}$ and $\sum_{\mathbf{i}\in\mathbf{I_j}} \rho_{\mathbf{j}}\left(x_{\mathbf{i}|\mathbf{j}}\right) = 1$. The primitive in our framework is the collection of all observed $\rho_{\mathbf{j}}$, $\rho = (\rho_{\mathbf{j}})_{\mathbf{j}\in\boldsymbol{J}}$. We call this collection a *dynamic stochastic choice function.*

Given ρ, we can define a Dynamic Random Utility Model (DRUM). Let U^t denote the set of all injective, monotone on $>^t$ (i.e., if $S, S' \in 2^{X^t} \setminus \{\emptyset\}$ and $S >^t S'$, then $\max_{s\in S} u^t(s) > \max_{s\in S'} u^t(s)$) utility functions that map X^t to $\mathbb{R}$. Also let $\mathcal{U} = \times_{t\in\mathcal{T}} U^t$ and $u = (u^t)_{t\in\mathcal{T}}$ be an element of it.

Definition 1 (DRUM)**.** The dynamic stochastic choice function ρ is consistent with DRUM if there exists a probability measure over $\mathcal{U}$, μ, such that

$$\rho_{\mathbf{j}}\left(x_{\mathbf{i}|\mathbf{j}}\right) = \int \prod_{t\in\mathcal{T}} \mathbb{1}\left(\operatorname*{arg\,max}_{y\in B^t_{j_t}} u^t(y) = x^t_{i_t|j_t}\right) d\mu(u)$$

for all $\mathbf{i} \in \mathbf{I_j}$ and $\mathbf{j} \in \mathbf{J}$.

When $T = 1$, DRUM coincides with RUM, where every agent maximizes her utility function u^1 over a menu and the analyst observes the distribution of consumers' choices. DRUM extends RUM by introducing a time dimension with unrestricted preference correlation across time. The stochastic utility process is captured by μ. Similar to RUM, DRUM does not restrict preference heterogeneity in cross-sections (i.e. across agents) and requires μ not to depend on the menu paths and the alternatives in the consumption space. Samuelson-Afriat's framework, in contrast to RUM and DRUM, does not use variation in choices of agents in cross-sections (i.e., it is directed to the individual level data or time series of choices). Thus, it does not restrict preferences of individuals in cross-sections. However, Samuelson-Afriat's framework, in contrast to DRUM, imposes a strict restriction that preferences are perfectly correlated across time (i.e. $u^t = u^s \ \mu-$a.s. for all $t, s \in \mathcal{T}$). We formalize these connections between RUM, Afriat's framework, and DRUM in Section 5.

Some examples of datasets where a dynamic stochastic choice function is (partially) observed are: (i) household longitudinal survey datasets; (ii) scanner datasets; and (iii) experimental datasets with panels of choice. In survey datasets (e.g., *Encuesta de Presupuestos Familiares* in Spain and Progresa Household Survey in Mexico, Deb *et al.*, 2021, Aguiar and Kashaev, 2021), information about household purchases is usually collected several times a year. For a given time period, budget variation across households is driven by spatial or regional price variation (Aguiar and Kashaev, 2021). Scanner datasets (e.g., Nielsen homescan data, Gauthier, 2018) contain information about weekly purchases of consumers. Budget variation in this case is driven by price variation across stores in each time period (Gauthier, 2021). In experimental settings, subjects often face at random few budget paths drawn from a common set of budgets (e.g., experiments on preferences over giving as in Porter and Adams, 2016). In our empirical application, we have a panel of choices over different menus of lotteries from an experimental data set collected by ABKK, (see also McCausland, Davis-Stober, Marley, Park and Brown, 2020 for a panel of choices in discrete choice.[8])

2.1. Preview of the Results: Binary Menus Example

We illustrate the setting and our main result with an example. We start with the static RUM setup for a choice set $X^t = \{x, y, z\}$, where one only observes binary menus. The order $>^t$ is assumed to be empty. Following McFadden-Richter we describe RUM as a finite mixture of deterministic types captured by a matrix A^t displayed in Table 1. Each column of A^t corresponds to a deterministic rational type r^t_i. (e.g., r^t_1 represents a strict rational order over X^t such that $xr^t_1yr^t_1z$). Note that there is some utility function u^t_1 such that $xr^t_1yr^t_1z$ if and only if $u^t_1(x) > u^t_1(y) > u^t_1(z)$. Each row corresponds to a choice from a binary menu

[8] In that paper they study static RUM in an individual setup using the time-series to estimate individual stochastic choice.

for each rational order. An element of A^t that corresponds to type r_i^t and pair $x', \{x', y'\}$ is equal to 1 if $x' r_i^t y'$ and zero otherwise. Now consider two periods where the choice set remains

	r_1^t	r_2^t	r_3^t	r_4^t	r_5^t	r_6^t
$x, \{x, y\}$	1	1	-	-	1	-
$y, \{x, y\}$	-	-	1	1	-	1
$x, \{x, z\}$	1	1	1	-	-	-
$z, \{x, z\}$	-	-	-	1	1	1
$y, \{y, z\}$	1	-	1	1	-	-
$z, \{y, z\}$	-	1	-	-	1	1

Table 1 – The matrix A^t for binary menus. “ $-$ ” corresponds to 0.

the same in time. DMs face (sequentially) two menus. Thus, menu paths are of the form $\{x', y'\}, \{x'', y''\}$ for all $x', y', x'', y'' \in X^t$. Given a menu path, DMs choose a choice path (e.g., the ordered tuple $(x, \{x, y\}; y, \{y, z\})$ indicating the choices from menu path $\{x, y\}, \{y, z\}$). Our primitive, or data set, is the collection of the joint probabilities of choice paths for each menu path, ρ. We show that ρ is consistent with DRUM if and only if (up to rearrangement) $\rho = (A^1 \otimes A^2)\nu$ for some vector $\nu \geq 0$ such that $\sum_i \nu_i = 1$, where $\otimes$ is the Kronecker product. The vector ν is a distribution over dynamic profiles of deterministic rational types or columns of $A^1 \otimes A^2$ (e.g., a dynamic preference profile (r_1^1, r_6^2) is such that preferences change from $x r_1^1 y r_1^1 z$ to $z r_6^2 y r_6^2 x$). This representation of DRUM as a mixture of deterministic dynamic rational types is called the $\mathcal{V}$-representation. The recursive structure of the $\mathcal{V}$-representation makes DRUM modular with the consequent computational gains. In addition, when $\mathcal{T} = \{1\}$ consistency of ρ with RUM is equivalent to the triangle condition

$$\rho_{\{x',y'\}}(x') + \rho_{\{y',z'\}}(y') - \rho_{\{x',z'\}}(x') \geq 0$$

for all $x', y', z' \in X^t$.[9] The triangle conditions can be summarized in a matrix H^t as displayed in Table 2. The triangle conditions can then be stated as $H^t \rho \geq 0$ for the static case. This is called the $\mathcal{H}$-representation of RUM.

Our results say that one can use this characterization of RUM to obtain testable conditions for

[9]The triangle conditions are equivalent to RUM with binary menus when $|X^t| \leq 5$ (Dridi, 1980).

$x, \{x,y\}$	$y, \{x,y\}$	$x, \{x,z\}$	$z, \{x,z\}$	$y, \{y,z\}$	$z, \{y,z\}$
1	-	-1	-	1	-
-1	-	1	-	-	1
-	1	1	-	-1	-
-	-1	-	1	1	-
1	-	-	1	-	-1
-	1	-	-1	-	1

Table 2 – The matrix H^t for binary menus.“ $-$ ” corresponds to 0.

DRUM in the form of linear inequalities by computing $(H^1 \otimes H^2)\rho \geq 0$. (We later show that, with more work, we can also obtain the full characterization of DRUM in this environment.) We call these conditions the dynamic triangle conditions. A dynamic triangle condition can be expressed recursively in the binary setup as

$$\begin{aligned} \mathrm{D}^{\triangle,1}_{z,x,y}((z,\{x,z\};x,\{x,z\})) &= \mathrm{D}^{\triangle,2}_{x,z,y}((z,\{x,z\};x,\{x,z\}) \\ &+ \mathrm{D}^{\triangle,2}_{x,z,y}((x,\{x,y\};x,\{x,z\}) - \mathrm{D}^{\triangle,2}_{x,z,y}((z,\{y,z\};x,\{x,z\}) \geq 0, \end{aligned}$$

where an instance of the dynamic triangle condition is

$$\mathrm{D}^{\triangle,2}_{x,z,y}((z,\{x,z\};x,\{x,z\})) = \rho_{\{x,z\},\{x,z\}}(z,x) + \rho_{\{x,z\},\{y,z\}}(z,z) - \rho_{\{x,z\},\{x,y\}}(z,x) \geq 0.$$

This instance means that holding constant the choice in period 1, the triangle inequality holds for ρ for a triple of choices in period 2.

Thus, the dynamic triangle condition applies recursively the triangle inequality in the first period to the triangle inequality in the second period.

Our results use the recursive version of the Weyl-Minkowski theorem to show how this insight —deriving the $\mathcal{V}$- or $\mathcal{H}$-representation of DRUM from its one-time $\mathcal{V}$- or $\mathcal{H}$-representations—can be generalized to obtain the characterization of DRUM for any finite time window and any collection of menus (i.e, beyond binary menus), both discrete (e.g., Li, 2021, Chambers *et al.*, 2021) and continuous (e.g., the demand setup of KS).

2.2. Finite Abstract Setup

We consider a nonempty finite grand choice set X^t in each $t \in \mathcal{T}$ with an empty (hence acyclic) order $>^t$, and assume the observed menus in each $t \in \mathcal{T}$ are all possible subsets of X^t with cardinality at least 2. This setup is a generalization of Li (2021), which assumes that $|X^t| \leq 3$, and Chambers *et al.* (2021), which effectively assumes that $T = 2$.

2.3. Demand Setup

Let $X^* \subseteq \mathbb{R}^K_+$ be the consumption space with finite $K \geq 2$ goods.[10] In each $t \in \mathcal{T}$, there are $J^t < \infty$ distinct budgets

$$B^{*,t}_j = \left\{ y \in X^* \,:\, p'_{j,t} y = w_{j,t} \right\}, \quad j \in \mathcal{J}^t = \{1, \dots, J^t\},$$

where $p_{j,t} \in \mathbb{R}^K_{++}$ is the vector of prices and $w_{j,t} > 0$ is the expenditure level.

Similar to the general setup, $\mathbf{j} = (j_t)_{t \in \mathcal{T}}$, $j_t \in \mathcal{J}^t$ encodes budgets that were faced by DMs in different time periods, referred to as budget paths. Let $\mathbf{J}$ be a set of all observed budget paths.

For every $\mathbf{j} \in \mathbf{J}$, let $\mathrm{P}_{\mathbf{j}}$ be a probability measure on the set of all Borel measurable subsets of $\times_{t \in \mathcal{T}} X^*$. The primitive in the demand framework is the collection of all observed $\mathrm{P}_{\mathbf{j}}$, $\mathrm{P} = (\mathrm{P}_{\mathbf{j}})_{\mathbf{j} \in \boldsymbol{J}}$. We call this collection a *dynamic stochastic demand system*.

Given P, we can define a Dynamic Random Demand Model (DRDM). Let U^* denote the set of all continuous, strictly concave, and monotone utility functions that map X^* to $\mathbb{R}$; $\mathcal{U}^* = \times_{t \in \mathcal{T}} U^*$, and $u^* = (u^{*t})_{t \in \mathcal{T}}$.

Definition 2 (DRDM)**.** The dynamic stochastic demand P is consistent with DRUM if there

[10] $\mathbb{R}^K_+$ denotes the set of component-wise nonnegative elements of the K-dimensional Euclidean space $\mathbb{R}^K$.

exists a probability measure over $\mathcal{U}^*$, μ^*, such that

$$\mathrm{P}_{\mathbf{j}}\left(\left(O^t\right)_{t\in\mathcal{T}}\right) = \int \prod_{t\in\mathcal{T}} \mathbb{1}\left(\arg\max_{y\in B^t_{j_t}} u^{*t}(y) \in O^t\right) d\mu^*(u^*)$$

for all $\mathbf{j} \in \mathbf{J}$ and all Borel measurable $O^t \subseteq X^*$, $t \in \mathcal{T}$.

We next show that DRDM is empirically equivalent to DRUM provided we appropriately specify the choice set and primitive order. Monotonicity of the utility functions generates choices on the budget hyperplane. In the RUM demand setting, KS and Kawaguchi (2017) showed that to establish that P is consistent with DRUM all possible Borel sets do not need to be checked. Stochastic rationalizability by RUM only depends on the probability of certain regions of the budget hyperplanes called *patches*.

For any $t \in \mathcal{T}$ and $j \in \mathcal{J}^t$, let $\{x^t_{i|j}\}_{i\in\mathcal{I}^t_j}$, $\mathcal{I}^t_j = \{1, \dots, I^t_j\}$, denote a finite partition of $B^{*,t}_j$ (each element of the partition is indexed by i).

Definition 3 (Patches)**.** For every $t \in \mathcal{T}$, let $\bigcup_{j\in\mathcal{J}^t}\{x^t_{i|j}\}$ be the coarsest partition of $\bigcup_{j\in\mathcal{J}^t} B^{*,t}_j$ such that

$$x^t_{i|j} \bigcap B^{*,t}_{j'} \in \{x^t_{i|j}, \emptyset\}$$

for any $j, j' \in \mathcal{J}^t$ and $i \in \mathcal{I}^t_j$. A set $x^t_{i|j}$ is called a patch. If $x^t_{i|j} \subseteq B^{*,t}_{j'}$ for some i and $j \neq j'$, then $x^t_{i|j}$ is called an intersection patch.

By definition, patches can only be strictly above, strictly below, or on budget hyperplanes. A typical patch belongs to one budget hyperplane. However, intersection patches always belong to several budget hyperplanes. The case for one time period, $K = 2$ goods and $J^t = 2$ budgets is depicted in Figure 1. Note that by definition $\{x^t_{i|j}\}$ is a partition of $B^{*,t}_j$ and I^t_j is the number of patches that form budget $B^{*,t}_j$.

The (discretized) choice set is

$$X^t = \bigcup_{i_t\in I^t_j, j\in\mathcal{J}^t} \{x_{i_t|j_t}\}.$$

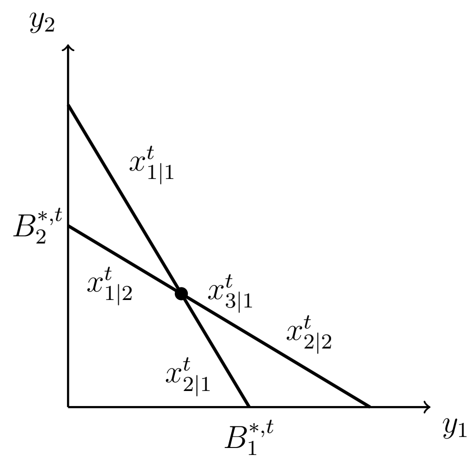

Figure 1 – Patches for the case with $K = 2$ goods and $J^t = 2$ budgets. The only intersection patch is $x^t_{3|1}$, which is the intersection of B^t_1 and B^t_2.

The primitive order $>^t$ is given by $S' >^t S$ for $S, S' \in 2^{X^t} \setminus \{\emptyset\}$ whenever for any $x_{i|j} \in S$ and $y \in x_{i|j}$ there exist $x_{i'|j'} \in S'$ and $y' \in x_{i'|j'}$ such that $y' > y$, where $>$ is the strict vector order on X^*. We define a menu as the collection of patches from the same budget hyperplane

$$B^t_{j_t} = \left\{x_{i_t|j_t}\right\}_{i_t \in I^t_j}.$$

Henceforth, we refer to a menu or budget interchangeably.

Let

$$\rho\left(x_{\mathbf{i}|\mathbf{j}}\right) = \mathrm{P}_{\mathbf{j}}\left(x_{\mathbf{i}|\mathbf{j}}\right)$$

denote the fraction of agents who pick from a choice path $x_{\mathbf{i}|\mathbf{j}}$ given a budget path $\mathbf{j}$.

The main building block of our demand framework is the dynamic stochastic choice function

$$\rho = \left(\rho\left(x_{\mathbf{i}|\mathbf{j}}\right)\right)_{\mathbf{j} \in \mathbf{J}, \mathbf{i} \in \mathbf{I}_{\mathbf{j}}}.$$

The vector ρ represents the distribution over finitely many patches and contains all the necessary information needed to determine whether P is consistent with DRDM and, in this discretized setup, consistent with DRUM.

Lemma 1. *The following are equivalent:*

(i) P is consistent with DRDM.

(ii) ρ is consistent with DRUM.

The proof of Lemma 1 follows from Kitamura and Stoye (2018) and Kawaguchi (2017). Next we provide some parametric examples of DRUM in this domain.

Example 1 (Dynamic Random Cobb-Douglas Utility)**.** Let $K = 2$ and $u^t(y_1, y_2) = y_1^{\alpha_t} y_2^{(1-\alpha_t)}$. The utility parameter α_t is random and such that $\alpha_t = \max\{\min\{\alpha_{t-1}+\epsilon_t, 1\}, 0\}$, where $(\epsilon_t)_{t\in\mathcal{T}}$ are independent and identically distributed mean-zero random innovations with variance σ^2. The dynamic stochastic demand generated by this utility function is consistent with DRUM as long as $(\alpha_t)_{t\in\mathcal{T}}$ is independent of prices and income.

Example 2 (Adams *et al.*, 2015)**.** For a deterministic utility $v : X^* \to \mathbb{R}$, the random utility at time $t \in \mathcal{T}$ is given by $u^t(x) = v(x) + \alpha_t' x$, where α_t is the random vector supported on $\mathbb{R}^K$. The dynamic stochastic demand generated by this utility function is consistent with DRUM if α_t is independent of prices and income.

In the two examples above, as well as in the examples in Section 3, we maintain the assumption that the distribution of preferences does not depend on prices and income. This assumption is satisfied in experimental setups such as the one in Porter and Adams (2016), McCausland *et al.* (2020), and ABKK. That said, it may not be realistic in other setups such as when saving is possible. This exogeneity assumption is relaxed in Section 2.4.

2.4. Endogenous Expenditure in the Demand Setup

In the demand setup we assumed that budgets are exogenously given. Here, we relax the exogeneity assumption by extending the results of Deb *et al.* (2021) to our setup. Our new model will cover the classical consumption smoothing problem with income uncertainty (Browning, 1989). Similarly to DRDM, we can define a Dynamic Random Augmented Demand Model (DRADM). Let V denote the set of all continuous, strictly concave, and monotone augmented utility functions that map $X^* \times \mathbb{R}_-$ to $\mathbb{R}$ and $\mathcal{V} = \times_{t\in\mathcal{T}} V$ be the Cartesian product of T repetitions of V.

Definition 4 (DRADM)**.** A dynamic stochastic demand P is consistent with DRADM if there exists a probability measure over $\mathcal{V}$, η, such that

$$\mathrm{P}_{\mathbf{j}}\left(\left(O^t\right)_{t\in\mathcal{T}}\right) = \int \prod_{t\in\mathcal{T}} \mathbb{1}\left(\arg\max_{y\in X^*} v^t(y, -p'_{j,t}y) \in O^t\right) d\eta(v)$$

for all $\mathbf{j} \in \mathbf{J}$ and all Borel measurable $O^t \subseteq X^*$, $t \in \mathcal{T}$, where $v = (v^t)_{t\in\mathcal{T}}$.

While DRDM is an extension of RUM to a dynamic setting (i.e., DRDM and RUM coincide when $T = 1$), DRADM is a dynamic extension of the Random Augmented Utility Model of Deb *et al.* (2021).

Example 3 (Consumption Smoothing with Income Uncertainty)**.** Consider a consumer with random income stream $y = (y_t)_{t\in\mathcal{T}}$ who maximizes the expected flow of instantaneous, concave, locally nonsatiated, and continuous utilities, u, given the budget constraints, discount factor δ, history of incomes captured by the information set I_t, and initial level of savings s_0. That is, at every time period τ the consumer solves

$$\max_{\{c_\tau(\cdot), s_\tau(\cdot)\}_{\tau=t,\dots,T}} \mathbb{E}\left[\sum_{\tau=t}^{T} \delta^{\tau-t} u(c_\tau(y)) \middle| I_\tau\right]$$

subject to

$$p'_\tau c_\tau(y) + s_t(y) = y_\tau + (1 + r_\tau)s_{\tau-1}(y).$$

The sequences of consumption and saving (policy) functions $(c_t(\cdot))_{t\in\mathcal{T}}$ and $(s_t(\cdot))_{t\in\mathcal{T}}$ fully describe the consumption and saving decisions of the consumer. In addition, we restrict these functions to depend only on the income history. That is, for all t, $c_t(y') = c_t(y)$ and

$s_t(y') = s_t(y)$ for all y and y' such that $y'_\tau = y_\tau$ for all $\tau \leq t$. The Bellman equation for this problem is

$$W_{t-1}(s_{t-1}) = \max_c \left[u(c) + \delta \mathbb{E} \left[W_t(y_t + (1+r_t)s_{t-1} - p'_t c) \middle| I_t \right] \right],$$

where W_t is the value function at time period t. Thus, one can define the state-dependent utility function as

$$\hat{v}^t(x, s_{t-1}) = u(x) + \delta \mathbb{E} \left[W_t(y_t + (1+r_t)s_{t-1}(y) - p'_t c) \middle| I_t \right].$$

Correlation in income across time would generate correlation between $\{\hat{v}^t\}_{t\in\mathcal{T}}$. One can define the augmented utility function as

$$v^t(x, -p'x) = u(x) + \delta \mathbb{E} \left[W_t(y_t + (1+r_t)s_{t-1}(y) - p'_t c) \middle| I_t \right].$$

Notice that the utility $\hat{v}^t$ depends on s_{t-1} only through the contemporaneous expenditure $p'x$. Correlation in income across time would generate correlation between $\{v^t\}_{t\in\mathcal{T}}$. If one assumes that different individuals have different u, δ, and y such that their joint distribution does not depend on prices, then this setup is a particular case of DRADM.

In Example 3, the random augmented utility stochastic process is independent of prices because prices are determined exogenously by supply and demand forces. Next, we characterize DRADM by using the fact that consistency with DRADM is equivalent to consistency with DRDM for a normalized budget path. A normalized budget path has the same price path $(p_{j,t})_{t\in\mathcal{T}}$ and income equal to 1. Using these normalized budgets, we can define patches as before to obtain

$$\rho\left(x_{\mathbf{i}|\mathbf{j}}\right) = \mathrm{P}_{\mathbf{j}}\left(\left\{y^t \in X^* \,:\, y^t/p'_{j,t}y^t \in x^t_{i|j}\right\}_{t\in\mathcal{T}}\right),$$

for all $\mathbf{i} \in \mathbf{I_j}$, $\mathbf{j} \in \mathbf{J}$. As before the stochastic choice function that corresponds to P is

$$\rho = \left(\rho(x_{\mathbf{i},\mathbf{j}})\right)_{\mathbf{i}\in\mathbf{I_j},\mathbf{j}\in\mathbf{J}}.$$

Similarly to DRDM, we rule out intersection patches. The choice set X^t and the partial order $>^t$ are defined analogously than in DRDM.

Lemma 2. *The following are equivalent:*

(i) P *is consistent with DRADM.*

(ii) ρ *is consistent with DRUM.*

The proof of Lemma 2 is omitted because it is analogous to the results in Deb *et al.* (2021).

3. Characterization of DRUM

Here we provide a characterization of rationalizability by DRUM when ρ is observed (estimable). The main result in this section is an analogue of the McFadden-Richter's and KS's results for RUM. Given the finite choice set, let a preference profile be $\mathbf{r} = (r^t)_{\in\mathcal{T}}$, where r^t is a linear order defined on the finite set of alternatives available at time t, X^t. We restrict these linear orders to be extensions of $>^t$ (i.e., for $S, S' \in 2^{X^t} \setminus \{\emptyset\}$ and $S >^t S'$, there is some $x \in S$ such that xr^ty for all $y \in S'$). Recall that $\mathbf{i}$ encodes choices in each time period. Given $\mathbf{r}$, we can encode choices in different time periods and menus in a vector $a_{\mathbf{r}}$ as

$$a_{\mathbf{r}} = \left(a_{\mathbf{r},\mathbf{i},\mathbf{j}}\right)_{\mathbf{j}\in\mathbf{J},\mathbf{i}\in\mathbf{I_j}},$$

with $a_{\mathbf{r},\mathbf{i},\mathbf{j}} = 1$ if the alternative $x^t_{i_t|j_t}$ is the best item available in $B^t_{j_t}$ according to r_t for all $t \in \mathcal{T}$ and $a_{\mathbf{r},\mathbf{i},\mathbf{j}} = 0$ otherwise. Denote $\mathcal{R}^t$ is the set of (strict) rational preferences in a given time period $t \in \mathcal{T}$. The set of dynamic rational preference profiles $\mathcal{R}$ is the set of all preference profile $\mathbf{r}$ for which there exists $u_r = (u_r^t)_{t\in\mathcal{T}} \in \mathcal{U}$ such that

$$a_{\mathbf{r},\mathbf{i},\mathbf{j}} = 1 \quad \iff \quad \forall t \in \mathcal{T},\ \arg\max_{x\in B_{j_t}^t} u_r^t(x) = x_{i_t|j_t}.$$

We form the matrix A_T by stacking the column vectors $a_{\mathbf{r}}$ for all preference profiles $\mathbf{r} \in \mathcal{R}$. The dimension of this matrix is $d_\rho \times |\mathcal{R}|$, where d_ρ is the length of vector ρ. This matrix will be used to provide a characterization of DRUM that is amenable to statistical testing.

The next axiom is the analogue of the McFadden-Richter's axiom for (static) stochastic revealed preferences (Border, 2007) and will provide a different characterization of DRUM.

Definition 5 (Axiom of Dynamic Stochastic Revealed Preference, ADSRP)**.** A stochastic choice function ρ satisfies ADSRP if for every finite sequence of pairs of menu and choice paths (including repetitions), k, $\{(\mathbf{i}_k, \mathbf{j}_k)\}$ such that $\mathbf{j}_k \in \mathbf{J}$ and $\mathbf{i}_k \in \mathbf{I}_{\mathbf{j}_k}$

$$\sum_k \rho\left(x_{\mathbf{i}_k|\mathbf{j}_k}\right) \le \max_{\mathbf{r}\in\mathcal{R}} \sum_k a_{\mathbf{r},\mathbf{i}_k,\mathbf{j}_k}.$$

The next theorem provides a full characterization of DRUM. Let

$$\Delta^L = \left\{ y \in \mathbb{R}_+^{L+1} \ :\ \sum_{l=1}^{L+1} y_l = 1 \right\}$$

denote the L-dimensional simplex.

Theorem 1. *The following are equivalent:*

(i) ρ is consistent with DRUM.

(ii) There exists $\nu \in \Delta^{|\mathcal{R}|-1}$ such that $\rho = A_T\nu$.

(iii) There exists $\nu \in \mathbb{R}_+^{|\mathcal{R}|}$ such that $\rho = A_T\nu$.

(iv) ρ satisfies ADSRP.

The proof of Theorem 1 is analogous to the proofs for RUM in McFadden and Richter (1990), McFadden (2005), KS, and Kawaguchi (2017). Theorem 1 (iii) is amenable to statistical testing using the test developed in KS. However, the number of columns in A_T grows exponentially with T. Thus, naively, testing DRUM may seem impossible for relatively small T even if one uses the tools of Smeulders, Cherchye and De Rock (2021). The next lemma shows that the computational complexity of computing A_T, $T \geq 1$ does not grow that much relatively to the computation complexity of computing A_1.

Lemma 3. *Let A^t be matrix constructed under the assumption that $\mathcal{T} = \{t\}$. That is, A^t is the matrix encoding static rational types at time t. Then $A_T = \otimes_{t \in \mathcal{T}} A^t$ up to permutation of its rows.*

Proof. Note that the k-th and the l-th columns of A_1 and A_2, a_k^1 and a_l^2, encode the choices of particular types of consumers at time $t = 1$ and $t = 2$ (i.e., their choices in each menu at $t = 1$ and $t = 2$). Since there are no restrictions across t on these deterministic types, we can generate the (k, l)-type, $a_k^1 \otimes a_l^2$ that encodes what is picked in pairs of menus where each menu is taken from two different time periods. Next, if we take some column from A^3 we can repeat the above step and obtain a composite type for three time periods. Repeating this exercise T times for all possible combinations of columns will lead to a matrix that is equal to A_T up to a permutation of rows. ■

Lemma 3 substantially simplifies the computation of A_T given that one can use the methods in KS and Smeulders *et al.* (2021) to construct A^t. In instances where the menu structure is such that $A^t = A^s$ for $t \neq s$ significant computational savings are achieved. Note that $A^t = A^s$ can occur without menus in t and s being the same. In fact, in the demand setting, A^t depends only on the intersection structure induced by the budgets and not on the specific prices (see examples in the next section). Lemma 3 also allows exploiting sparsity because the Kronecker product propagates any zero entry in A_t. The Kronecker product structure also illustrates that DRUM is modular because the structure of A_T is built from its static components. This property allows one to parallelize the computation of A_T.[11] This modularity is exploited to obtain a recursive characterization of DRUM.

Unfortunately, the DRUM characterization in Theorem 1 does not provide an intuitive understanding of the behavioral implications of DRUM. In the next sections, we provide an intuitive characterization of DRUM. This characterization demonstrate that DRUM provides additional implications relative to RUM in longitudinal data. That is, we show that requiring consistency with (static) RUM for all conditional and marginal probabilities is not enough. In fact, the new conditions will affect the joint distribution ρ.

4. Behavioral Characterization of DRUM via Linear Inequality Restrictions

In order to understand the behavioral implications of DRUM, we provide a characterization of many of its special cases via linear inequalities. We also provide a way to obtain a general characterization of DRUM via linear inequalities when its static counterpart is known. First, we need some preliminary mathematical results.

[11] Since each factor of the Kronecker product can be computed independently, we can parallelize along the time dimension.

$\mathcal{H}$ and $\mathcal{V}$-representations

Theorem 1 (iii) states that to test whether ρ is consistent with DRUM it is enough to check whether it belongs to a convex cone

$$\{A_T v \,:\, v \geq 0\}\,.$$

This is called the $\mathcal{V}$-representation of the cone. The Weyl-Minkowski theorem states that there exists an equivalent representation of it (the $\mathcal{H}$-representation) via some matrix B_T:

$$\{z \,:\, H_T z \geq 0\}\,.$$

The $\mathcal{V}$-representation states that the observed distribution over choices is a finite mixture of deterministic types (Kitamura and Stoye, 2018, Smeulders *et al.*, 2021).[12] Unfortunately, the $\mathcal{V}$-representation does not give any direct restrictions on the observed ρ. As a result, the analyst can hardly infer any helpful intuition about the empirical content of DRUM using the $\mathcal{V}$-representation. In contrast, the $\mathcal{H}$-representation can deliver direct and intuitive restrictions on the data (see Section 4.1 or the dynamic BM inequalities in Section 4.2).

In theory, if one possesses A_T, one can obtain H_T. Unfortunately, the construction of H_T from A_T is a nontrivial task that becomes computationally burdensome (Kitamura and Stoye, 2018) even for moderate T since the number of columns of A_T grows exponentially with T. In Lemma 3, we showed that one could use the recursive structure of A_T to simplify its construction substantially. In this section, we show that the same intuition carries over to the construction of H_T: one can move from the $\mathcal{H}$-representation of RUM to the $\mathcal{H}$-representation of DRUM with small computational costs. Our next result generalizes the Weyl-Minkowski theorem in a direction that is useful for our recursive setup. Henceforth, we assume that all

[12] KS were the first to notice that in the static case checking if a stochastic demand is consistent with RUM amounts to checking if its vector representation belongs to a convex cone. They also introduced the Weyl-Minkowski theorem to the study of RUM in economics.

cones are finite dimensional and are subsets of a Euclidean vector space.

Proposition 1. If

$$\left\{K^t v \,:\, v \geq 0\right\} = \left\{z \,:\, L^t z \geq 0\right\}$$

for all $t \in \mathcal{T}$, then

$$\left\{\left(\otimes_{t\in\mathcal{T}} K^t\right) v \,:\, v \geq 0\right\} \subseteq \left\{z \,:\, \left(\otimes_{t\in\mathcal{T}} L^t\right) z \geq 0\right\}.$$

Proposition 1 is a direct extension of Theorem A in Aubrun *et al.* (2021) and Theorem 7.15 in de Bruyn (2020) to more than 2 time periods.[13] Note that if K^t represents a model that can be expressed as mixtures of deterministic behavior (i.e., columns of K^t), then Proposition 1 allows one to easily construct testable conditions of the dynamic extensions of this model using its one-time $\mathcal{H}$-representations.

Next, we provide not only testable conditions but a full equivalence of the recursive $\mathcal{H}$-representation via Kronecker products and the corresponding $\mathcal{V}$-representation. Define the Kronecker power for a matrix C, $C^{\otimes_k} = \otimes_{j=1}^{k} C$ for any integer $k \geq 1$.[14] We say that $\{Cv \,:\, v \geq 0\}$ is proper if C is full row rank, the cone is closed, and any line in the vector space that contains the cone is not in the cone. For any ϕ^t in the interior of $\{L^{t\prime} v \,:\, v \geq 0\}$ (e.g. ϕ^t is a strict convex combination of columns of $L^{t\prime}$) and any $k \geq 1$ define the projection map

$$\gamma_k^{\phi^t} = \frac{1}{k}\sum_{j=1}^{k} \phi^{t,\otimes(j-1)} \otimes I^t \otimes \phi^{t,\otimes(k-j)},$$

where I^t is the identity matrix in the vector space containing the cone $\{K^t v \,:\, v \geq 0\}$. Define

$$\Gamma_{\mathbf{k}}^{\boldsymbol{\phi}} = I^1 \otimes \left(\otimes_{t\in\mathcal{T}\setminus\{1\}} \gamma_{k_t}^{\phi^t}\right)$$

for a given collection of static operators $\gamma_{k_t}^{\phi^t}$, $t \in \mathcal{T} \setminus \{1\}$.

[13] We thank Chris Chambers for pointing out an error on the proof of a previous version of this result.

[14] For $k = 0$, the Kronecker power is equal to scalar 1.

Theorem 2. *Suppose K^t is proper for all $t \in \mathcal{T}$. Then*

$$\left\{\left(\otimes_{t\in\mathcal{T}}K^t\right)v \,:\, v \geq 0\right\} = \bigcap_{k_1,\cdots,k_T\geq 1}\left\{\Gamma_{\mathbf{k}}^{\phi\prime}z \,:\, \left(\otimes_{t\in\mathcal{T}}L^{t,\otimes_{k_t}}\right)z \geq 0\right\}.$$

Moreover, K^t does not have full column rank for at most one $t \in \mathcal{T}$ if and only if

$$\left\{\left(\otimes_{t\in\mathcal{T}}K^t\right)v \,:\, v \geq 0\right\} = \left\{z \,:\, \left(\otimes_{t\in\mathcal{T}}L^t\right)z \geq 0\right\}.$$

For $T = 2$, Theorem 2 is proved in Aubrun *et al.* (2022). It easily extends to more than two time periods due to the associativity of the Kronecker product.[15] The moreover part of Theorem 2 is established in Theorem A in Aubrun *et al.* (2021) and Corollary 4. The first part of the statement essentially provides an approximation result that allows us to obtain the $\mathcal{H}$-representation of a dynamic model via its one time counterparts by using extensions of the model to k_t time periods. For some cones the number of extensions can be infinite but in practice we can use a finite number of extensions with the knowledge that in the limit this produces an exact characterization. The simplest case of the previous result happens under additional full column rank requirements.

$\mathcal{H}$-representation of DRUM

Proposition 1 gives necessary conditions for building the $\mathcal{H}$-representation of DRUM from its static components (i.e. $K^t = A^t$ and $L^t = H^t$, where H^t is the matrix from the $\mathcal{H}$-representation of a cone generated by A^t). We show that despite the fact that A^t does not generate a proper cone (it is never of full row rank), we can use Theorem 2 to obtain sufficient conditions. The row rank is not full because of the "adding-up" constraint—one alternative has to be picked from *every* menu. Hence, the sum of all rows belonging to the same menu will give the row of ones. In the running example with binary menus, with matrix A^t given

[15] See also Remark 1 on page 9 of Aubrun *et al.* (2022).

by Table 1, the sum of the first two rows is equal to the sum of the third and the fourth rows and is equal to the sum of the last two rows. However, Theorem 2 can still be used to obtain the characterization of DRUM as the next theorem demonstrates. To formalize this, consider the following submatrix of A^t, $t \in \mathcal{T}$: from every menu except the first one, pick the last alternative and remove the corresponding row from A^t. Let A^{t*} denote the resulting matrix. Observe that when $\mathcal{T} = \{t\}$, ρ is consistent with DRUM if and only ρ^*, defined analogously to A^{t*}, is such that $A^{t*}\nu^* = \rho^*$ for some $\nu^* \geq 0$. In particular, given the simplex constraints on a given ρ, we can safely drop adding-up constraints in matrix H^t from now on when computing the $\mathcal{H}$-representation of A^t. The reason is that adding-up constraints are guaranteed to hold. Before stating the next theorem we need to introduce a key behavioral condition implied by DRUM.

Definition 6 (Stability)**.** We say that ρ is stable if $\sum_{i \in \mathcal{I}_j^t} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right)$ is the same for all $j \in \mathcal{J}^t$ for any $t \in \mathcal{T}$ and $x_{\mathbf{i}|\mathbf{j}}$.

Stability means that the marginal distribution of choices at any t does not depend on the menu in any other $t' \neq t$. Under stability, the marginal distribution of choices will not change due to the menu the consumers faced in the past or the menus the consumers will face in the future. Recall, we have assumed that the stochastic utility process does not depend on the budgets. This condition is an implication of that assumption.[16] Stability was first defined in Strzalecki (2021). Chambers *et al.* (2021) call this condition in their domain marginality.

We also need a notion of uniqueness of RUM and DRUM that is directly connected to the requirement of full column rank in Theorem 2.

Definition 7 (Uniqueness)**.** We say that A^t generates a unique RUM when the system $\rho = A^t\nu$ has a unique solution for all stochastic choice functions ρ. Also, we say DRUM (associated with matrix $\otimes_{t \in \mathcal{T}} A^t$) satisfies uniqueness when for all $t \in \mathcal{T} \setminus \{t'\}$, A^t generates a unique RUM, for a given period t'.

[16]Stability and the simplex constraints on ρ are formally defining a vector subspace within which A^t is associated with a proper cone.

Set $\phi^{*,t}$ to be the average of all columns of $H^{t\prime}$. In other words, $\phi^{*,t}$ represents a testable linear inequality of static RUM such that for $\mathcal{T} = \{t\}$, $\phi^{*,t\prime}\rho \geq 0$. Define as well implicitly the associated operator $\Gamma_{\mathbf{k}}^{\phi^{*}\prime}$.

Theorem 3. *Assume that A^{t*} is full row rank for all $t \in \mathcal{T}$. Then ρ is consistent with DRUM if and only if ρ is stable and*

$$\rho \in \bigcap_{k_1,\cdots,k_T \geq 1} \left\{\Gamma_{\mathbf{k}}^{\phi^{*}\prime} z \,:\, \left(\otimes_{t\in\mathcal{T}} H^{t,\otimes k_t}\right) z \geq 0\right\}. \tag{1}$$

Moreover, ρ is consistent with unique DRUM if and only if ρ is stable and $(\otimes_{t\in\mathcal{T}} H^{t})\rho \geq 0$.

The condition that A^{t*} is full row rank for all $t \in \mathcal{T}$ is satisfied in all the examples we are aware of. Importantly, it holds for the abstract setup of Li (2021), Chambers *et al.* (2021) as proved by Dogan and Yildiz (2022). It holds as well in the finite abstract setup with limited menu variation as proven in Saito (2017). We conjecture this condition to be true in many other settings and verified it in the demand setup with up to 6 budgets per period with maximal intersection pattern, with $K \in \{2, 3, 4, 5\}$.[17]

Note that stability is a set of equality restrictions on ρ. Since any equality restriction can be represented as two inequality restrictions, Theorem 3 allows us to obtain the $\mathcal{H}$-representation of DRUM from the $\mathcal{H}$-representation of its static components recursively for any time window, for the unique DRUM case. In other words, one just needs to derive the $\mathcal{H}$-representation of (unique) RUM and then convert it to the dynamic setting and add the constraints implied by stability. This delivers a substantial gain over the direct computation of the $\mathcal{H}$-representation since the existing numerical algorithms transforming $\mathcal{V}$-representations to $\mathcal{H}$-representations are known to work for small and moderate-size problems only. That is, computational complexity of the dynamic problem is only bounded by the computational complexity of the static one.

[17] Note that the fact that the full row rank condition is the only condition that needs to be verified in each case as the cone associated with RUM is closed, and any line in the vector space that contains the cone is not in the cone.

Remarkably, if DRUM is not unique, then there is no hope that the $\mathcal{H}$-representation of DRUM is just a Kronecker product of the static ones. Indeed, Theorem 2 provides a necessary and sufficient conditions for this equivalence to hold. This result also shows that the problem of providing a full characterization of DRUM is a hard problem and that is why it has remained an open question until our contribution. For the case, of nonunique DRUM (e.g., more than two periods without uniqueness in the static factors), we still obtain the recursive linear inequalities, implied by the Kronecker product of the static RUM inequalities, as necessary conditions of consistency with DRUM. But, to obtain sufficient condition we have to do more work. Yet, our results provides an explicit way to compute the additional or emergent restrictions on ρ consistent with DRUM. Indeed, applying the results in Aubrun *et al.* (2022), we obtain condition (1) that describes a decreasing sequence of outer approximations to the convex cone associated with DRUM. Doherty, Parrilo and Spedalieri (2004) shows that using these outer approximations can do a good job approximating some cones of interest for finite k. Note as well that the convex cone associated with DRUM is finitely generated, and so there are finitely many linear inequalities characterizing DRUM regardless of the value of k_t. The general characterization of (static) RUM for our demand setup via $\mathcal{H}$-representation is yet to be discovered (Stoye, 2019). Only special cases are fully solved: the case of 2 budgets (Hoderlein and Stoye, 2014), and the case of 3 goods and 3 budgets (Kitamura and Stoye, 2018). This stands in contrast with the abstract setup–without monotonicity–and full menu variation solved in Block and Marschak (1960) and Falmagne (1978). Fortunately, the BM inequalities can be modified in our discretized setup, as we will see below, to deal with the demand setup. Nevertheless, our result implies that once the generic $\mathcal{H}$-representation of RUM becomes available, the analogous DRUM characterization will also become available.

Back to the Binary Menus Example.

In our running example, A^t is not full column rank. That means that the dynamic triangle conditions (and stability) are necessary but not sufficient conditions for DRUM. In our running example, for $T = 2$, we will obtain explicitly all components of equation (1), and explain how they provide testable implications for DRUM that become sufficient as we take large enough k. For $k = 1$, we conclude that if ρ is consistent with DRUM then it satisfies $\otimes_{t\in\mathcal{T}} H^t\rho \geq 0$ or the dynamic triangle conditions. Then, we focus on the second extension of the theory for $k = 2$. This means that we consider 3 virtual periods, $T^{v,3} = 3$. Let $\mathbf{J}^v$ be the set of all menu paths in the virtual time window. Define ρ^v as a dynamic stochastic choice function on $\mathbf{J}^v$.

To simplify the exposition and to obtain reduction of dimensionality, we will use the fact that A^t, in Table 1, can be simplified without loss of generality by computing A^{t*} as the submatrix with all rows of A^t except for row 4 and row 6. Due to the simplex constraints of DRUM these rows are redundant. Then, we can obtain the $\mathcal{H}$-representation matrix H^{t*} from Table 2 by deleting columns 4 and 6, and rows 2, 4, 5, 6. We have to retain the nonnegativity constraints that are not associated to the deleted rows in A^t. Since we are imposing the simplex and stability constraints on ρ, we can simplify equation (1) to work with the reduced H^{t*} and the reduced ρ^*, the subarray of ρ obtained after we delete all entries with a choice path containing one of the deleted rows of A^t. (We define $\rho^{*,v}$ in an analogous way to ρ^*.)

We set $\phi^{*,2}$ as the average of all triangle conditions in Table 2 and nonnegativity constraints:

$$\phi^{*,2} = \frac{1}{6}\begin{pmatrix} 2 & 2 & 1 & 1 \end{pmatrix}'.$$

Recall I^t is a diagonal matrix of dimension 4 for all $t \in \mathcal{T}$, so we will denote it by I. The projection mapping $\gamma_2^{\phi^{*,2}}$ is given by a matrix of size 4×16 whose rows with nonnegative entries add up to 1:

$$\begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{12} & \frac{1}{12} & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{12} & 0 & 0 & 0 & \frac{1}{12} & 0 & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{12} & \frac{1}{12} & 0 & \frac{1}{12} & 0 & 0 & 0 & \frac{1}{12} & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} & 0 & 0 & \frac{1}{12} & 0 \\ 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & \frac{1}{12} & \frac{1}{6} & \frac{1}{6} & \frac{1}{12} & \frac{1}{6} \end{pmatrix}.$$

The matrix $\gamma_2^{\phi^{*,2}}$ can be interpreted as a linear operator taking weighted averages. In fact, the associated matrix $\Gamma_2^{\phi^{*,2}\prime}$ reduces any ρ^v of size 64×1 to a vector of size 16×1, that is the size of ρ^*. Recall that we defined the matrix H^t by Table 2 and nonnegativity constraints. If ρ is consistent with DRUM, then there is some virtual ρ^v such that the following will be satisfied:

$$\begin{aligned} &\rho^* = \Gamma_2^{\phi^{*,2}\prime} \rho^{*,v} \\ &\otimes_{t=1}^3 H^{*,t} \rho^{*,v} \geq 0. \end{aligned}$$

It should be evident that the dynamic triangle inequality conditions are implied by the previous conditions, but also new emergent conditions appear. In particular, we write down explicitly the first entries of ρ^* vector (with each entry representing the probability of a choice path) in terms of the $\rho^{*,v}$ (the remaining entries can be computed easily by the reader):

$$\begin{aligned} \rho_1^* &= \frac{\rho_1^{*,v}}{3} + \frac{\rho_2^{*,v}}{6} + \frac{\rho_3^{*,v}}{12} + \frac{\rho_4^{*,v}}{12} + \frac{\rho_5^{*,v}}{6} + \frac{\rho_9^{*,v}}{12} + \frac{\rho_{13}^{*,v}}{12} \\ \rho_2^* &= \frac{\rho_2^{*,v}}{6} + \frac{\rho_5^{*,v}}{6} + \frac{\rho_6^{*,v}}{3} + \frac{\rho_7^{*,v}}{12} + \frac{\rho_8^{*,v}}{12} + \frac{\rho_{10}^{*,v}}{12} + \frac{\rho_{14}^{*,v}}{12}. \end{aligned}$$

These two equations illustrate that ρ_1 and ρ_2 become connected or dependent through $\rho^{*,v}$ entries $\rho_2^{*,v}$ and $\rho_5^{*,v}$. In addition, we can obtain new inequalities explicitly from these relations. Behaviorally, these additional conditions say that when ρ is consistent with DRUM, then ρ must be the result of projecting back to $T = 2$, a ρ^v that is consistent with the dynamic triangle conditions, in a larger time window. The projection mapping $\Gamma_2^{\phi^{*,2}\prime}$ can be interpreted as a *measuring device* based on the average of testable implications of the static case $\phi^{*,2}$. This measurement device asks what the choices of the virtual DMs associated with ρ^v in the actual time window are.[18] We can make analogous statements for any $k \geq 3$.

4.1. The Simple-Setup: 2 budgets per time period

Here we illustrate our main results in the demand environment with two budgets in each time period $B_1^{*,t}$ and $B_2^{*,t}$ such that $B_1^{*,t} \cap B_2^{*,t} \neq \emptyset$ and $w_{1,t}/p_{1,t,K} > w_{2,t}/p_{2,t,K}$ for all $t \in \mathcal{T}$. To simplify the analysis, we assume that the intersection patches are picked with probability zero. Thus, in each time period there are four patches $x^t_{1|1}, x^t_{2|1}, x^t_{1|2}$, and $x^t_{2|2}$ (see Figure 2 for a graphical representation of the case with $K = 2$ goods).[19] We call choice path configurations implied by these 4 patches the *simple-setup* choice paths. An example of a budget path for $T = 2$ is $(2, 1)$ (i.e. B_2^1 and B_1^2), an example of a choice path in this budget path is $\left(x^1_{1|2}, x^2_{1|1}\right)$. In this setup, there are 3 rational demand types per time period that are described in Table 3.[20] Each demand type $\theta^t_{i,j}$ picks i-th patch in menu B_1^t and j-th patch in menu B_2^t at time t.

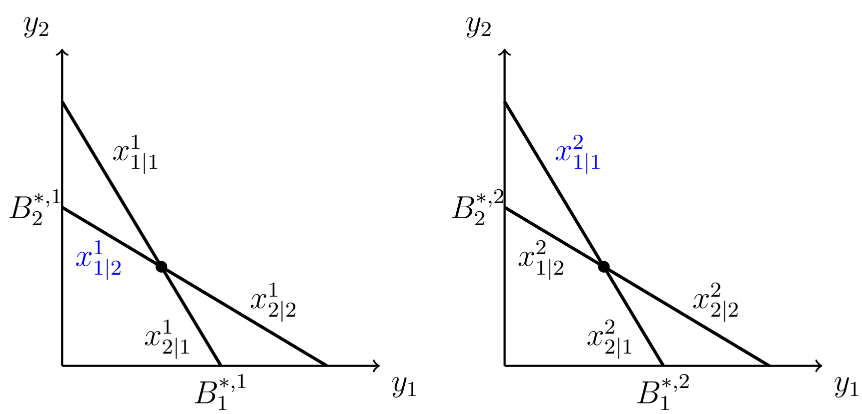

Figure 2 – Simple-setup for $K = 2$ goods and no intersection patches.

[18]These emergent conditions are related in their mathematical structure to the entanglement phenomenon in quantum physics (Aubrun *et al.*, 2021, 2022).

[19]Formally, $x^t_{1|1} = \{y \in B_1^{*,t} : p'_{2,t}y > w_{2,t}\}$, $x^t_{2|1} = \{y \in B_1^{*,t} : p'_{2,t}y < w_{2,t}\}$, $x^t_{1|2} = \{y \in B_2^t : p'_{1,t}y < w_{1,t}\}$, and $x^t_{2|2} = \{y \in B_2^{*,t} : p'_{1,t}y > w_{1,t}\}$.

[20]We use the convenient notation developed in Im and Rehbeck (2021).

Type/Budget	B_1^t	B_2^t
$\theta_{1,1}^t$	$x_{1\|1}^t$	$x_{1\|2}^t$
$\theta_{1,2}^t$	$x_{1\|1}^t$	$x_{2\|2}^t$
$\theta_{2,2}^t$	$x_{2\|1}^t$	$x_{2\|2}^t$

Table 3 – Choices of 3 rational types in menus B_1^t and B_2^t at time t.

	$\theta_{1,1}^t$	$\theta_{1,2}^t$	$\theta_{2,2}^t$
$x_{1\|1}^t$	1	1	-
$x_{2\|1}^t$	-	-	1
$x_{1\|2}^t$	1	-	-
$x_{2\|2}^t$	-	1	1

Table 4 – The matrix A^t for 2 budgets. "$-$" corresponds to 0.

We can now write down the associated A_T matrix. Since there are two intersecting budgets in every time period, $A^t = A^{t'}$ for all $t, t' \in \mathcal{T}$. Thus, by Lemma 3, we can compute the matrix A^t for one period. We display the matrix A^t in Table 4. Note that it is easy to verify that A^t has full column rank, thus DRUM associated with $\otimes_{t\in\mathcal{T}} A^t$ is unique according to our definition. This allows us to obtain necessary and sufficient conditions explicitly. Using A^t, we can write down the matrix A_T for any T (e.g., See Table 5 for $T = 2$ or $A_T = A^1 \otimes A^2$).

D-**monotonicity.–** We introduce a new behavioral restriction on ρ that together with stability characterize DRUM in the simple-setup. We first introduce a static notion of dominance among patches.

Definition 8 (Patch-Revealed Dominance)**.** We say that patch $x_{i|j}^t$ is revealed dominant to $x_{i'|j'}^t$ if $x_{i_t|j_t}^t >^t x_{i'_t|j'_t}^t$.

Patch-revealed dominance requires that all elements in the dominant patch are directly revealed preferred (in the Afriat's sense) to the dominated patch, and that all the elements of the dominated patch are not directly revealed preferred to the elements of the dominant patch. We can visualize this ordering in Figure 2, where $x_{1|1}^1 >^t x_{1|2}^1$ and $x_{2|2}^2 >^t x_{2|1}^2$. Let $x_{\mathbf{i}|\mathbf{j}} \downarrow_t x_{i'_t|j'_t}^t$ denote a choice path where the t-th component of $x_{\mathbf{i}|\mathbf{j}}$ was replaced by $x_{i'_t|j'_t}^t$. We

	$(\theta^1_{1,1},\theta^2_{1,1})$	$(\theta^1_{1,1},\theta^2_{1,2})$	$(\theta^1_{1,1},\theta^2_{2,2})$	$(\theta^1_{1,2},\theta^2_{1,1})$	$(\theta^1_{1,2},\theta^2_{1,2})$	$(\theta^1_{1,2},\theta^2_{2,2})$	$(\theta^1_{2,2},\theta^2_{1,1})$	$(\theta^1_{2,2},\theta^2_{1,2})$	$(\theta^1_{2,2},\theta^2_{2,2})$
$\left(x^1_{1\|1},x^2_{1\|1}\right)$	1	1	-	1	1	-	-	-	-
$\left(x^1_{1\|1},x^2_{2\|1}\right)$	-	-	1	-	-	1	-	-	-
$\left(x^1_{1\|1},x^2_{1\|2}\right)$	1	-	-	1	-	-	-	-	-
$\left(x^1_{1\|1},x^2_{2\|2}\right)$	-	1	1	-	1	1	-	-	-
$\left(x^1_{2\|1},x^2_{1\|1}\right)$	-	-	-	-	-	-	1	1	-
$\left(x^1_{2\|1},x^2_{2\|1}\right)$	-	-	-	-	-	-	-	-	1
$\left(x^1_{2\|1},x^2_{1\|2}\right)$	-	-	-	-	-	-	1	-	-
$\left(x^1_{2\|1},x^2_{2\|2}\right)$	-	-	-	-	-	-	-	1	1
$\left(x^1_{1\|2},x^2_{1\|1}\right)$	1	1	-	-	-	-	-	-	-
$\left(x^1_{1\|2},x^2_{2\|1}\right)$	-	-	1	-	-	-	-	-	-
$\left(x^1_{1\|2},x^2_{1\|2}\right)$	1	-	-	-	-	-	-	-	-
$\left(x^1_{1\|2},x^2_{2\|2}\right)$	-	1	1	-	-	-	-	-	-
$\left(x^1_{2\|2},x^2_{1\|1}\right)$	-	-	-	1	1	-	1	1	-
$\left(x^1_{2\|2},x^2_{2\|1}\right)$	-	-	-	-	-	1	-	-	1
$\left(x^1_{2\|2},x^2_{1\|2}\right)$	-	-	-	1	-	-	1	-	-
$\left(x^1_{2\|2},x^2_{2\|2}\right)$	-	-	-	-	1	1	-	1	1

Table 5 – The matrix A_T for 2 time periods with 2 budgets per period. "$-$" corresponds to 0.

show that if ρ is consistent with DRUM and $x^t_{i'_t|j'_t} >^t x^t_{i_t|j_t}$, then

$$\rho\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^t_{i'_t|j'_t}\right) \geq \rho\left(x_{\mathbf{i}|\mathbf{j}}\right).$$

We illustrate the necessity for the simple case where $T = 1$, then $A_T\nu = \rho$ can be rewritten as

$$\nu_1 + \nu_2 = \rho(x^1_{1|1}), \qquad \nu_3 = \rho(x^1_{2|1}),$$
$$\nu_1 = \rho(x^1_{1|2}), \qquad \nu_2 + \nu_3 = \rho(x^1_{2|2}).$$

Then the following two inequalities, consistent with monotonicity, have to be satisfied

$$0 \leq \nu_2 = \rho(x^1_{1|1}) - \rho(x^1_{1|2}),$$
$$0 \leq \nu_2 = \rho(x^1_{2|2}) - \rho(x^1_{2|1}).$$

In fact, we can write down the $\mathcal{H}$-representation of static RUM, for the simple setup using matrix $H^{2,t}$ defined in Table 6, capturing this monotonicity condition such that $H^{2,t}\rho \geq 0$.

For $T \geq 2$, DRUM implies that ρ satisfies dynamic monotonicity. For illustrative purposes, set $T = 2$. Then we get the new condition by exploiting the recursive structure of the matrix

$x^t_{1\|1}$	$x^t_{2\|1}$	$x^t_{1\|2}$	$x^t_{2\|2}$
1	-	-1	-
1	-	-	-
-	1	-	-
-	-	1	-

Table 6 – The $\mathcal{H}$-representation of RUM for 2 goods and 2 budgets.

A_T :

$$\rho = A_T\nu = A^1 \otimes A^2\nu = \begin{pmatrix} A^1 & A^1 & 0 \\ 0 & 0 & A^1 \\ A^1 & 0 & 0 \\ 0 & A^1 & A^1 \end{pmatrix} \begin{pmatrix} \nu^1_1 \\ \nu^1_2 \\ \nu^1_3 \end{pmatrix} = \begin{pmatrix} A^1(\nu^1_1 + \nu^1_2) \\ A^1\nu^1_3 \\ A^1\nu^1_1 \\ A^1(\nu^1_2 + \nu^1_3) \end{pmatrix}.$$

We can derive the following system of equations

$$A^1\nu_1 = \left[\rho^1_{1|1} - \rho^1_{1|2}\right],$$
$$A^1\nu_2 = \left[\rho^1_{2|2} - \rho^1_{2|1}\right],$$

where $\rho^1_{i|j}$ is a vector of all probabilities that correspond to all choice paths that contain patch $x^1_{i|j}$. For example,

$$\rho^1_{1|1} = \begin{pmatrix} \rho\left(\left(x^1_{1|1}, x^2_{1|1}\right)\right) \\ \rho\left(\left(x^1_{1|1}, x^2_{2|1}\right)\right) \\ \rho\left(\left(x^1_{1|1}, x^2_{1|2}\right)\right) \\ \rho\left(\left(x^1_{1|1}, x^2_{2|2}\right)\right). \end{pmatrix}$$

Repeating the argument for $T = 1$, from $A^1\nu_1 = \left[\rho^1_{1|1} - \rho^1_{1|2}\right]$, we derive that

$$0 \le \left[\rho\left(\left(x^1_{1|1}, x^2_{1|1}\right)\right) - \rho\left(\left(x^1_{1|2}, x^2_{1|1}\right)\right)\right] - \left[\rho\left(\left(x^1_{1|1}, x^2_{1|2}\right)\right) - \rho\left(\left(x^1_{1|2}, x^2_{1|2}\right)\right)\right]. \tag{2}$$

The following inequalities have to hold under DRUM

$$\rho\left(\left(x^1_{1|1}, x^2_{1|1}\right)\right) - \rho\left(\left(x^1_{1|2}, x^2_{1|1}\right)\right) \ge 0,$$

$$\rho\left(\left(x^1_{1|1}, x^2_{1|2}\right)\right) - \rho\left(\left(x^1_{1|2}, x^2_{1|2}\right)\right) \geq 0.$$

Thus, Inequality (2) imposes a restriction on how the distribution over patches can grow. In particular, it implies that the increase in probability caused by switching from patch $x^1_{1|2}$ to the dominant patch $x^1_{1|1}$ is bigger if the patch in the second time period, $x^2_{1|1}$, dominates $x^2_{1|2}$. In other words, there is some form of complementarity between dominant patches in different time periods. The above arguments can be generalized for an arbitrary but finite time window. However, we need to work with higher-order differences that are in fact the Kronecker product of the static monotonicity conditions. Next, we introduce the difference operator.

Definition 9 (Difference operator)**.** For any t, $x^t_{i'_t|i'_t}$, and, $x_{\mathbf{i}|\mathbf{j}}$, let $\mathrm{D}\left(x^t_{i'_t|j'_t}\right)[\cdot]$ be a linear operator such that

$$\mathrm{D}\left(x^t_{i'_t|j'_t}\right)\left[f(x_{\mathbf{i}|\mathbf{j}})\right] = f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^t_{i'_t|j'_t}\right) - f\left(x_{\mathbf{i}|\mathbf{j}}\right)$$

for any f that maps choice paths to reals.

The D operator applied to ρ calculates the difference in ρ when only one patch in a choice path was replaced. When the operator is applied twice to two different time periods, it computes the difference in differences. That is, for $t_1 \neq t_2$

$$\begin{aligned}
&\mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\mathrm{D}\left(x^{t_1}_{i'_{t_1}|j'_{t_1}}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] = \mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_{t_1} x^{t_1}_{i'_{t_1}|j'_{t_1}}\right) - f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] = \\
&\mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_{t_1} x^{t_1}_{i'_{t_1}|j'_{t_1}}\right)\right] - \mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] = \\
&\left[f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_{t_1} x^{t_1}_{i'_{t_1}|j'_{t_1}} \downarrow_{t_2} x^{t_2}_{i'_{t_2}|j'_{t_2}}\right) - f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_{t_1} x^{t_1}_{i'_{t_1}|j'_{t_1}}\right)\right] - \left[f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_{t_2} x^{t_2}_{i'_{t_2}|j'_{t_2}}\right) - f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right],
\end{aligned}$$

where the second equality uses linearity of $\mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)$.

Similarly, we can apply D any K number of times. Let

$$\boldsymbol{\mathcal{T}} = \left\{(t_k)_{k=1}^K \,:\, t_k \in \mathcal{T},\ t_k < t_{k+1}, K \leq T\right\}$$

be a collection of all possible increasing sequences of length at most T. For any $\mathbf{t} \in \boldsymbol{\mathcal{T}}$ and any $x^{\mathbf{t}}_{\mathbf{i}'|\mathbf{j}'} = \left(x^t_{i'_t|j'_t}\right)_{t\in\mathbf{t}}$ define

$$\mathrm{D}\left(x^{\mathbf{t}}_{\mathbf{i}'|\mathbf{j}'}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] = \mathrm{D}\left(x^{t_K}_{i'_{t_K}|j'_{t_K}}\right)\dots\mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\mathrm{D}\left(x^{t_1}_{i'_{t_1}|j'_{t_1}}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right],$$

where K is the length of $\mathbf{t}$.

Definition 10 (D-monotonicity)**.** We say that ρ is D-monotone if for any $\mathbf{t} \in \boldsymbol{\mathcal{T}}$, $x^{\mathbf{t}}_{\mathbf{i}'|\mathbf{j}'}$, and any $x_{\mathbf{i}|\mathbf{j}}$ such that $x^t_{i'_t|j'_t} >^t x^t_{i_t|j_t}$ for all $t \in \mathbf{t}$

$$\mathrm{D}\left(x^{\mathbf{t}}_{\mathbf{i}'|\mathbf{j}'}\right)\left[\rho\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] \geq 0.$$

D-monotonicity is the generalization of the Weak Axiom of Stochastic Revealed Preference (WASRP) introduced by Bandyopadhyay *et al.* (1999) to our dynamic setup. Since the set of utility functions we consider is monotone, our condition coincides with the *stochastic substitutability* condition in Bandyopadhyay, Dasgupta and Pattanaik (2004) when $\mathcal{T} = \{t\}$. If the domain of choices is incomplete, Dasgupta and Pattanaik (2007) shows that WASRP is sufficient but not necessary for regularity. This observation translates to the dynamic case as well.[21] We emphasize that D-monotonicity imposes stronger restrictions on the data than WASRP. Also, we note that the strength of these restrictions increases as the time window expands (see our Monte Carlo experiments). As such, D-monotonicity can be used to derive informative counterfactual bounds on the longitudinal distribution of out-of-sample demand with sufficiently long panels. Given these observations, we are ready to present the main result of this section.

Theorem 4. *For the simple-setup, the following are equivalent:*

(i) ρ is consistent with DRUM.

[21] In that regard, D-monotonicity is not implied by any of the conditions derived in Li (2021) or Chambers *et al.* (2021) that require complete menu variation and use generalizations of the static regularity conditions for the dynamic or correlated case.

	$x^2_{1\|1}$	$x^2_{2\|1}$	$x^2_{1\|2}$	$x^2_{2\|2}$
$x^1_{1\|1}$	3/4	-	3/4	-
$x^1_{2\|1}$	-	1/4	1/4	-
$x^1_{1\|2}$	-	1/4	1/4	
$x^1_{2\|2}$	3/4	-	3/4	-

Table 7 – Matrix representation of ρ for $T = 2$ that violates D-monotonicity, but satisfies simple stability. "-" corresponds to 0

(ii) ρ *is* D*-monotone and stable.*

(iii) ρ *is stable and* $\otimes_{t\in\mathcal{T}} H^{2,t}\rho \geq 0$.

The fact that (i) implies (ii) is easy to verify. The converse is proved constructively. The equivalence of (iii) and (i) is a corollary of Theorem 3. Importantly, in that case, we can set $k_t = 1$ for all $t \in \mathcal{T}$. This is because A^t in the simple-setup is full column rank and we can use Theorem 2.

Corollary 1. *For the simple-setup if* $\rho = A_T\nu = A_T\nu'$ *for some* $\nu, \nu' \in \Delta^{|\mathcal{R}-1|}$, *then* $\nu = \nu'$.

Stability and D-monotonicity are logically independent as we demonstrate in the next counterexample of DRUM.[22]

Example 4 (Violation of D-monotonicity)**.** Consider the stochastic demand presented in Table 7. It satisfies stability. However, it fails to satisfy D-monotonicity because $\rho\left(\left(x^1_{1|2}, x^2_{1|1}\right)\right) - \rho\left(\left(x^1_{1|2}, x^2_{1|2}\right)\right) = -\frac{1}{4}$ and $x^2_{1|1} >^t x^2_{1|2}$.

Generalization of D**-monotonicity for the demand setup with** 3 **goods and** 3 **budgets per period.–** Our results can be used to construct a set of necessary and sufficient conditions for DRUM in the demand setup on the basis of the $\mathcal{H}$-representation of (static, $\mathcal{T} = \{t\}$) DRDM for the case of 3 goods and 3 budgets.[23] The $\mathcal{V}$-representation in this case is given by matrix A^t in Table 8. The $\mathcal{H}$-representation, $H^{3,t}$, is displayed in Table 9 (without the nonnegativity

[22]Another example of a ρ that fails both conditions of the simple setup is discussed in Section 5.
[23]In each period there are 3 budgets with maximal intersections as in Example 3.2 in KS.

1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	$x^t_{1\|1}$
-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	-	-	-	-	-	-	-	-	$x^t_{2\|1}$
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	-	-	-	$x^t_{3\|1}$
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	$x^t_{4\|1}$
1	1	1	1	-	-	-	-	-	-	-	-	1	1	1	1	-	1	1	-	-	-	1	1	-	$x^t_{1\|2}$
-	-	-	-	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	$x^t_{2\|2}$
-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	1	-	-	-	1	-	-	-	1	$x^t_{3\|2}$
-	-	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	1	-	-	-	$x^t_{4\|2}$
1	-	-	-	1	-	-	-	1	-	1	-	1	-	-	-	1	1	-	1	1	1	1	-	1	$x^t_{1\|3}$
-	1	-	-	-	1	-	-	-	-	-	-	-	1	-	-	-	-	1	-	-	-	-	1	-	$x^t_{2\|3}$
-	-	1	-	-	-	1	-	-	1	-	1	-	-	1	-	-	-	-	-	-	-	-	-	-	$x^t_{3\|3}$
-	-	-	1	-	-	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	$x^t_{4\|3}$

Table 8 – A^t for 3 goods and 3 budgets. "-" corresponds to 0.

constraints). We have that $H^{3,t} = H^{3,s}$ for any $s, t \in \mathcal{T}$. Then we can establish the following direct implication of Theorem 3 since A^t in this case is such that A^{t*} is full row rank.

$x^t_{1\|1}$	$x^t_{2\|1}$	$x^t_{3\|1}$	$x^t_{4\|1}$	$x^t_{1\|2}$	$x^t_{2\|2}$	$x^t_{3\|2}$	$x^t_{4\|2}$	$x^t_{1\|3}$	$x^t_{2\|3}$	$x^t_{3\|3}$	$x^t_{4\|3}$
-	-	-	-1	-	-	-	-1	1	1	1	-
-	-	-	-1	1	-	-	-	1	-	-	-
1	-	-	-	1	-	-	-	-	-	-	-1
1	-	-	-	-	-	-	-1	1	-	-	-
-	-1	-	-1	1	-	1	-	-	-	-	-
-	-	-	-	1	1	-	-	-	-1	-	-1
-	-	-1	-1	-	-	-	-	1	1	-	-

Table 9 – The $\mathcal{H}$-representation of RUM for 3 goods and 3 budgets excluding nonnegativity.

Corollary 2. *For the demand setup ($K = 3$, $J^t = 3$ for all $t \in \mathcal{T}$), the following are equivalent:*

(i) ρ is consistent with DRUM.

(ii) ρ is stable and

$$\rho \in \bigcap_{k_1, \cdots, k_T \geq 1} \left\{ \Gamma_{\mathbf{k}}^{\phi^*\prime} z \,:\, \left(\otimes_{t \in \mathcal{T}} H^{3,t,\otimes_{k_t}} \right) z \geq 0 \right\}.$$

For $k = 1$ we obtain necessary conditions in the form of $(\otimes_{t \in \mathcal{T}} H^{3,t})\rho \geq 0$. The last 3 rows of the matrix displayed in Table 9 and their Kronecker product in time are consistent with

D-monotonicity.[24] The rest of the conditions produce new testable implications. For instance, for $T = 2$, the following conditions are implied by monotonicity in row 4:

$$D^{*1}(x^2_{i|j}) = \left[\rho\left(\left(x^1_{1|2}, x^2_{i|j}\right)\right) + \rho\left(\left(x^1_{3|2}, x^2_{i|j}\right)\right) - \rho\left(\left(x^1_{2|1}, x^2_{i|j}\right)\right) - \rho\left(\left(x^1_{4|1}, x^2_{i|j}\right)\right)\right] \geq 0.$$

The interaction of row 3, a triangle condition, and row 4, a monotonicity condition, gives the implication:

$$D^{*1}(x^2_{1|1}) + D^{*1}(x^2_{1|2}) - D^{*1}(x^2_{4|3}) \geq 0.$$

Obtaining sufficient conditions in terms of a $H^{3,t}$ matrix requires more work because A^t is no longer full column rank in this case. Nevertheless, we can obtain sufficient conditions explicitly only from the knowledge of $H^{3,t}$. We set $\phi^{*,t}$ as the average of all conditions implied by $H^{3,t}$. The rest of the computations for any $k \geq 2$ can be done in the same way as in the running example for binary menus.

4.2. Characterization of DRUM via Recursive Block Marschak Inequalities

A BM-like characterization of RUM in the case of limited menu variation or a primitive order in the choice space (e.g., demand setup) did not exist. Here, we provide a characterization, via linear inequalities, based on the BM inequalities for the case of limited menu variability and with monotonicity of utilities.

Recall that $\mathcal{J}^t$ denotes the set of observed menus at time t. Let $\bar{\mathcal{J}}^t = \{1, 2, \ldots, 2^{|X^t|} - 1\}$ be the "virtual" set of menus such that there is a one-to-one mapping between $j_t \in \bar{\mathcal{J}}^t$ and a nonempty subset of X^t (i.e., the virtual data set has full menu variation). We also assume

[24] Note that if $\mathrm{P} = (\mathrm{P}_{\mathbf{j}})_{\mathbf{j}\in\mathbf{J}}$ is consistent with DRDM, then the stochastic demand system consisting of 2 different budget paths $(\mathrm{P}_{\mathbf{j}}, \mathrm{P}_{\mathbf{j}'})$ would also be consistent with DRDM. Moreover, note that $(\mathrm{P}_{\mathbf{j}}, \mathrm{P}_{\mathbf{j}'})$ form the simple-setup since at every time period there are exactly two budgets. Thus, $\mathbb{D}$-monotonicity is a simple necessary condition of DRUM.

that this mapping is such that the first $\mathcal{J}^t$ indexes correspond to observed menus B^t_j. That is $\mathcal{J}^t$ is the set of all observed menus and $\bar{\mathcal{J}}^t \setminus \mathcal{J}^t$ is the set of all "virtual menus" that were not observed in the data. Using this extended definition of menus, we can define a set of all (including the observed ones) menu paths $\bar{\mathbf{J}}$. Note that the set of observed menu paths $\mathbf{J}$ is a subset of $\bar{\mathbf{J}}$. Finally, as before, given a menu path $\mathbf{j} \in \bar{\mathbf{J}}$, let $\bar{\rho}$ be a distribution over choice paths $\mathbf{i}$ in $\mathbf{j}$. That is, $\bar{\rho}\left(x_{\mathbf{i}|\mathbf{j}}\right) \geq 0$ and $\sum_{\mathbf{i}\in\mathbf{I_j}} \bar{\rho}\left(x_{\mathbf{i}|\mathbf{j}}\right) = 1$ for all $\mathbf{j} \in \bar{\mathbf{J}}$. Recall that the observed ρ is defined as

$$\rho = \left(\rho\left(x_{\mathbf{i}|\mathbf{j}}\right)\right)_{\mathbf{j}\in\mathbf{J},\mathbf{i}\in\mathbf{I_j}}.$$

Similarly, the extended stochastic choice function is denoted by

$$\bar{\rho} = \left(\bar{\rho}\left(x_{\mathbf{i}|\mathbf{j}}\right)\right)_{\mathbf{j}\in\bar{\mathbf{J}},\mathbf{i}\in\mathbf{I_j}}.$$

Next we define some properties of $\bar{\rho}$.

Definition 11. We say that $\bar{\rho}$ *agrees* with ρ if they coincide on observed menu paths. That is, $\bar{\rho}\left(x_{\mathbf{i}|\mathbf{j}}\right) = \rho\left(x_{\mathbf{i}|\mathbf{j}}\right)$ for all $\mathbf{j} \in \mathbf{J}$ and $\mathbf{i} \in \mathbf{I_j}$.

If $\bar{\rho}$ agrees with ρ then it is an extension of the latter to virtual menu paths.

Definition 12 (Increasing Utility (IU) Consistency)**.** We say that $\bar{\rho}$ is IU-consistent if $\bar{\rho}\left(x_{\mathbf{i}|\mathbf{j}}\right) = 0$ whenever there exists $t \in \mathcal{T}$ and $I'_t \subseteq \mathcal{I}^t_{j_t}$ such that $\cup_{i'_t\in I'_t}\{x_{i'_t|j_t}\} >^t \{x_{i_t|j_t}\}$.

IU-consistency captures the empirical content of monotonicity of the utility functions with respect to $>^t$.

Definition 13 (BM inequalities)**.** We say that $\bar{\rho}$ satisfies the BM inequalities if for all $t \in \mathcal{T}$, $\mathbf{j} \in \bar{\mathbf{J}}$, and $\mathbf{i} \in \mathbf{I_j}$

$$\mathbb{B}^t(\mathbf{i},\mathbf{j}) = \sum_{j'_t : B^t_{j_t} \subseteq B^t_{j'_t}} (-1)^{\left|B^t_{j'_t} \setminus B^t_{j_t}\right|} \bar{\rho}\left(x_{\mathbf{i}|\mathbf{j}'}\right) \geq 0.$$

Note that the BM inequalities are linear in $\bar{\rho}$. Hence, we can construct matrix $\bar{H}^t$ with elements in $\{-1, 0, 1\}$ such that each row of $\bar{H}^t$ corresponds to one BM inequality.

We are ready to state the two main results of this section. First, a BM characterization of RUM for our setup and later an analogous characterization for DRUM.

Theorem 5. *Let $\mathcal{T} = \{t\}$. For a given ρ the following are equivalent:*

(i) ρ is consistent with RUM.

(ii) There exists $\bar{\rho}$ that agrees with ρ, is IU-consistent, and satisfies the BM inequalities.

(iii) There exists $\bar{\rho}$ that agrees with ρ, is IU-consistent, and is such that $\bar{H}^t\bar{\rho} \geq 0$.

The BM inequalities provide the $\mathcal{H}$-representation of RUM for the static case. For this extended setup with $\bar{\rho}$ we say that $\bar{A}^t$ generates a unique RUM when the system $\bar{\rho} = \bar{A}^t\nu$ has a unique solution for all choice functions $\bar{\rho}$. Also, we say DRUM (associated to matrix $\otimes_{t\in\mathcal{T}}\bar{A}^t$) satisfies uniqueness when for all $t \in \mathcal{T} \setminus \{t'\}$, $\bar{A}^t$ generates a unique RUM, for an arbitrary period t'. We can obtain the modified $\mathcal{H}$-representation for unique RUM for $\mathcal{T} = \{t\}$ from Turansick (2022), $\bar{H}^t$, with rows corresponding to: (i) BM inequalities, (ii) nonnegativity constraints, and (iii) those obtained from the uniqueness restriction in Theorem 1 in Turansick (2022).[25]

Theorem 6. *The following are equivalent.*

(i) ρ is consistent with DRUM satisfying uniqueness.

(ii) ρ is IU-consistent, stable, and satisfies $(\otimes_{t\in\mathcal{T}}\bar{H}^t)\bar{\rho} \geq 0$.

The result is a direct consequence of Theorem 3 and Theorem 5. Theorem 6 is not exactly a $\mathcal{H}$-representation of DRUM. It becomes one when all menus in $\bar{\mathbf{J}}$ are observed like it is

[25]The conditions in Turansick (2022) are equivalent to some BM polynomials to be equal to zero. This can be expressed in our setup by adding to the $\mathcal{H}$-representation the row corresponding the relevant BM polynomial and the same row multiplied by -1. This guarantees that the relevant BM polynomial is zero.

assumed in Chambers *et al.* (2021), Li (2021). Note, moreover, that our proof can be used in the environments of Chambers *et al.* (2021) and Li (2021) in the finite abstract setting because the primitive order $>^t$ can be empty. Notice that the results in Li (2021) are a special case of our corollary because when the choice set is of cardinality 3 or less, RUM is unique (Fishburn, 1998, Turansick, 2022). Our result also generalizes Theorem 3 in Chambers *et al.* (2021) that is equivalent to the special case $T = 2$.

We generalize the BM inequalities for the case of unobserved menus. Even if for our primitive this recursive characterization of DRUM is not a $\mathcal{H}$-representation of DRUM, this characterization has several advantages: (i) it avoids the computation of matrix A_T associated with the $\mathcal{V}$-representation, which can be computationally burdensome (KS note computing A^t is NP hard); (ii) it provides further intuition about the additional empirical bite of DRUM in comparison to RUM; and (iii) it means DRUM can be tested with a linear program.[26]

We remark that one cannot weaken the uniqueness assumption at all. In fact, the result fails to be true when A^t is associated with non-unique static RUM for more than two periods. In that case, we have to go back to Theorem 7.

We set $\phi^{*,t}$ for any $t \in \mathcal{T}$ as the average of rows of $\bar{H}^t$. The typical entry of $\bar{H}^t$ corresponding to the BM inequalities is:

$$\bar{H}^t_{(i_t,j_t),(i'_t,j'_t)} = (-1)^{|B^t_{j'_t} \setminus B^t_{j_t}|} \mathbb{1}\left(x_{i_t|j_t} = x_{i'_t|j'_t}, B^t_{j_t} \subseteq B^t_{j'_t}, x_{i_t|j'_t} \in B^t_{j'_t}\right).$$

Hence, $\phi^{*,t}_{i_{t'},j_{t'}} = \sum_{i_t,j_t} \bar{H}^t_{(i_t,j_t),(i'_t,j'_t)}$. As a direct application of Theorem 3, we can establish the following result.

Theorem 7. *The following are equivalent.*

(i) ρ is consistent with DRUM.

[26] For the statistical problem, we can use tools in KS and Fang, Santos, Shaikh and Torgovitsky (2023).

(ii) There exists $\bar{\rho}$ that agrees with ρ, is IU-consistent, stable, and satisfies

$$\bar{\rho} \in \bigcap_{k_1,\cdots,k_T \geq 1} \left\{ \Gamma_{\mathbf{k}}^{\phi^*\prime} z \ : \ \left(\otimes_{t \in \mathcal{T}} \bar{H}^{t, \otimes_{k_t}} \right) z \geq 0 \right\}.$$

To understand the intuition behind the $\mathcal{H}$-representation of DRUM we focus on a necessary condition implied by it. We define a new set of inequalities we call DRUM-BM.

Definition 14 (DRUM-BM inequalities)**.** We say that $\bar{\rho}$ satisfies the DRUM-BM inequalities if for all $t \in \mathcal{T}$, $\mathbf{j} \in \bar{\mathbf{J}}$, and $\mathbf{i} \in \mathbf{I}_{\mathbf{j}}$, $\mathbb{B}_t(\mathbf{i}, \mathbf{j}) \geq 0$, where $\mathbb{B}_T(\mathbf{i}, \mathbf{j}) = \mathbb{B}^T(\mathbf{i}, \mathbf{j})$ and

$$\mathbb{B}_t(\mathbf{i}, \mathbf{j}) = \sum_{j'_t : B^t_{j_t} \subseteq B^t_{j'_t}} (-1)^{\left| B^t_{j'_t} \setminus B^t_{j_t} \right|} \mathbb{B}_{t+1}(\mathbf{i}, \mathbf{j}')$$

for all $t \in \mathcal{T} \setminus \{T\}$.

Now we establish the following result.

Corollary 3. *If ρ is consistent with DRUM, then $\bar{\rho}$ satisfies the DRUM-BM inequalities.*

The DRUM-BM inequalities are as intuitive as the BM inequalities for the static case but they are not enough for DRUM. This is because these conditions interact with the hierarchical theory extensions for the nonunique DRUM case, giving rise to emergent conditions.

5. Relationship with Samuelson-Afriat's and McFadden-Richter's frameworks

In this section, we study the implications of DRUM for time series and cross sections. First, we look at a time series with the assumption of constant utility across time periods as in Samuelson-Afriat's framework. In this case, DRUM implies that the (deterministic) Strong

Axiom of Revealed Preference (SARP) has to hold in time series. Second, we study cross-sections, as the ones described in McFadden and Richter (1990), McFadden (2005), that are obtained by *marginalizing* or *pooling* panels. Marginalization and pooling correspond to empirical practices of using cross-sections that correspond to one or many time periods, respectively. We show that if ρ is consistent with DRUM, then any marginal distribution derived from it is rationalizable by RUM. At the same time, not every DRUM consistent panel is RUM rationalizable when pooled. Importantly, marginal consistency with RUM is not sufficient for consistency with DRUM.

5.1. Samuelson-Afriat's framework

DRUM has no testable implications for a time series without further restrictions. That is, if we observe $\rho_{\mathbf{j}}$ for a single budget path $\mathbf{j}$, then there are no testable restrictions of DRUM. (We need at least two observed budget paths to test DRUM.) However, in Samuelson-Afriat's framework, one only needs time-series of choices from budgets to test utility maximization. The reason for this is that in Samuelson-Afriat's framework there is an additional assumption on the stochastic process, namely, that μ is such that $u^t = u^s$ $\mu-$a.s. for all $t, s \in \mathcal{T}$. We call this restriction *constancy* of the stochastic utility process. Under this restriction, the testable implications of DRUM in a time series are re-established. We need some preliminaries to formalize this intuition. To simplify the exposition all the results in this sections are for the demand setup.

Definition 15 (Strong Axiom of Revealed Path Dominance, SARPD)**.** For a given $\mathbf{j} \in \mathbf{J}$, $\rho_{\mathbf{j}} = (\rho(x_{\mathbf{i}|\mathbf{j}}))_{\mathbf{i}\in \mathbf{I}_{\mathbf{j}}}$ satisfies SARPD if

$$\rho\left(x_{\mathbf{i}|\mathbf{j}}\right) = 0$$

whenever there is a finite set of patches from $x_{\mathbf{i}|\mathbf{j}}$, $\left\{x^{t_n}_{i_{t_n}|j_{t_n}}\right\}_{n=1}^{N}$, such that $x^{t_1}_{i_{t_1}|j_{t_1}} \succeq^* x^{t_2}_{i_{t_2}|j_{t_2}} \succeq^* \cdots \succeq^* x^{t_N}_{i_{t_N}|j_{t_N}}$ and $x^{t_N}_{i_{t_N}|j_{t_N}} \succeq^* x^{t_1}_{i_{t_1}|j_{t_1}}$(where $x^{t}_{i_t|j_t} \succeq^* x^{s}_{i_s|j_s}$ whenever $x \in x^{t}_{i_t|j_t}$ and $y \in x^{s}_{i_s|j_s}$

and $p'_{j_t}(x-y) \geq 0$).

SARPD requires that the probability of observing a choice path that contains consumption bundles that form a revealed preference cycle is zero. It is analogous to the Strong Axiom of Revealed Preferences (SARP) in Samuelson-Afriat's framework. Using SARPD, we can establish the following result.

Proposition 2. If ρ is rationalized by DRUM with μ that satisfies constancy, then $\rho_{\mathbf{j}}$ satisfies SARPD for all $\mathbf{j} \in \mathbf{J}$.

Note that DRUM bounds above the probability of choice paths that contain a revealed preference cycle. To see this, we consider again the simple-setup with $T = 2$. There are two choice paths that contain a revealed preference cycle: $\left(x^1_{1|2}, x^2_{2|1}\right)$ and $\left(x^1_{2|1}, x^2_{1|2}\right)$. We focus on the first choice path without loss of generality. Using D-monotonicity we know that

$$\rho\left(\left(x^1_{1|2}, x^2_{2|1}\right)\right) \leq \rho\left(\left(x^1_{1|1}, x^2_{2|2}\right)\right).$$

This means that the probability of a choice path that contains a violation of SARP, or a revealed preference (RP) cycle, is bounded above by the probability of a choice path that contains no RP cycles. That is, DRUM restricts the probability of choice paths with RP cycles meaningfully. This *endogenous* bound on the probability of a choice path that contains a revealed preference cycle has an important advantage with respect to measures of deviations from rationality like the Critical Cost Efficiency Index (CCEI) (Afriat, 1973). Indeed, in that literature it is an open question how to set a threshold below which the level of deviations from static utility maximization is deemed reasonable. In our setup, we convert this problem into a population one and then bound endogenously the fraction of consumers that have choices that involve revealed preference cycles. Importantly, notice that if ρ is degenerate taking values on $\{0, 1\}$ for a given budget path then D-monotonicity is equivalent to the Weak Axiom of Revealed Preference by Samuelson (1938) in Samuelson-Afriat's framework. To see this, note that the probability of choice paths with RP cycles of size 2 (i.e., violations of WARP) under the degeneracy of ρ must be zero.

5.2. Marginal and Conditional Distributions

Given a budget path $\mathbf{j}$, let $\rho^{\mathrm{c}}_{t,\mathbf{j}}$ and $\rho^{\mathrm{m}}_{t,\mathbf{j}}$ be the conditional and the marginal distributions over patches implied by $\rho_{\mathbf{j}}$. That is,

$$\rho^{\mathrm{c}}_{t,\mathbf{j}}\left(x_{\mathbf{i}|\mathbf{j}}\right) = \frac{\rho\left(x_{\mathbf{i}|\mathbf{j}}\right)}{\sum_{i\in\mathcal{I}^t_{j_t}} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right)},$$

$$\rho^{\mathrm{m}}_{t,\mathbf{j}}\left(x_{i_t|j_t}\right) = \sum_{\tau\in\mathcal{T}\setminus\{t\}} \sum_{i\in\mathcal{I}^\tau_{j_\tau}} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right),$$

where the conditional distribution is defined only when $\sum_{i\in\mathcal{I}^t_{j_t}} \rho\left(x_{i_t|j_t}\right) \neq 0$. Given the marginal distribution of a budget path, we can also define the slicing distribution as

$$\rho^{\mathrm{s}}_t\left(x_{i_t|j_t}\right) = \sum_{\mathbf{j}\in\mathbf{J}} \rho^{\mathrm{m}}_{t,\mathbf{j}}\left(x_{i_t|j_t}\right) F(\mathbf{j}|j_t),$$

where $F(\mathbf{j}|j_t)$ is the conditional probability of observing the budget path $\mathbf{j}$ conditional on the t-th budget being j_t in the data. The slicing distribution is a mixture of marginal distributions. It corresponds to the situation where the researcher only focuses on one cross-section.

Proposition 3. If ρ is rationalized by DRUM, then $\rho^{\mathrm{c}}_{t,\mathbf{j}}$, $\rho^{\mathrm{m}}_{t,\mathbf{j}}$, and ρ^{s}_t are rationalized by RUM for any $t \in \mathcal{T}$ and $\mathbf{j} \in \mathbf{J}$.

Proposition 3 means that if ρ is consistent with DRUM then the data are consistent with RUM in any given cross-section (slice). In this sense, the empirical implications of DRUM when an analyst has access only to a slice of choices is the same as the empirical implications of RUM. However, consistency of the marginal or slicing distributions does not exhaust the empirical content of DRUM. This is illustrated in Example 5.

	$x^2_{1\|1}$	$x^2_{2\|1}$	$x^2_{1\|2}$	$x^2_{2\|2}$
$x^1_{1\|1}$	1/6	1/3	2/3	-
$x^1_{2\|1}$	1/3	1/6	1/6	1/6
$x^1_{1\|2}$	1/6	1/3	2/3	-
$x^1_{2\|2}$	1/3	1/6	1/6	1/6

Table 10 – Matrix representation of ρ that is consistent with RUM after slicing, but is not consistent with DRUM

Example 5. [Marginals are consistent with WASRP but not rationalized by DRUM] Consider ρ presented in Table 10. This ρ violates stability and D$-$monotonicity. So DRUM cannot possibly explain it. At the same time its marginal probabilities at $t=1$ are consistent with the WASRP: $\rho^{\mathrm{m}}_{1,(2,1)}\left(x^1_{1|2}\right)=\frac{1}{2}$, $\rho^{\mathrm{m}}_{1,(1,1)}\left(x^1_{2|1}\right)=\frac{1}{2}$; and $\rho^{\mathrm{m}}_{1,(2,2)}\left(x^1_{1|2}\right)=\frac{2}{3}$, $\rho^{\mathrm{m}}_{1,(1,2)}\left(x^1_{2|1}\right)=\frac{1}{3}$. Thus, each of these marginal distributions is consistent with RUM.[27] Moreover, the slicing distribution would satisfy $\rho^{\mathrm{s}}_1\left(x^1_{2|1}\right)=F((1,1)|1)\frac{1}{2}+F((1,2)|1)\frac{1}{3}$ and $\rho^{\mathrm{s}}_1\left(x^1_{1|2}\right)=F((2,1)|2)\frac{1}{2}+F((2,2)|2)\frac{2}{3}$. As a result, depending on F,

$$\rho^{\mathrm{s}}_1\left(x^1_{1|2}\right)+\rho^{\mathrm{s}}_1\left(x^1_{2|1}\right)\in[5/6,7/6]\,.$$

Thus, if, for example, all budget paths are observed with equal conditional probabilities, then $\rho^{\mathrm{s}}_1\left(x^1_{1|2}\right)+\rho^{\mathrm{s}}_1\left(x^1_{2|1}\right)=1$. Thus, the slicing distribution is also consistent with RUM.

5.3. Pooling

In practice, and in the absence of panel variation, several years or time periods of choices from budgets are pooled before testing for consistency with RUM (Kitamura and Stoye, 2018, Deb *et al.*, 2021). Here we explore a potential pitfall of this practice. We show that a panel dataset that is consistent with DRUM when pooled may not be consistent with RUM. The spurious rejection of rationality may be driven by the fact that pooling requires us to ignore

[27]Recall that WASRP is the necessary and sufficient condition for marginal probabilities to be rationalized by RUM in the sense of Proposition 3.

time labels and imposes the restriction that the distribution of preferences is independent across time.

First, we formally define *pooling*. To simplify the exposition, assume that $B_j^{*,t} \neq B_{j'}^{*,t'}$ for all $t, t' \in \mathcal{T}$, $j \in \mathcal{J}^t$, and $j' \in \mathcal{J}^{t'}$. That is, there are no repeated budgets across time and agents. Let $\mathcal{J} = \{1, 2, \ldots, J\}$, where $J = \sum_{t \in \mathcal{T}} J^t$ is the total number of budgets.

Definition 16 (Pooled Patches)**.** Let

$$\mathcal{X} = \bigcup_{t \in \mathcal{T}} \bigcup_{j \in \mathcal{J}^t} \{\xi_{k|j}^t\}$$

be the coarsest partition of $\bigcup_{t \in \mathcal{T}} \bigcup_{j \in \mathcal{J}^t} B_j^{*,t}$ such that

$$\xi_{k|j}^t \bigcap B_{j'}^{*,t} \in \{\xi_{k|j}^t, \emptyset\}$$

for any j, j' and k.

The pooled patches $\{\xi_{k|j}^t\}$ partition every $x_{i|j}^t$ since $B_j^{*,t}$ may now intersect with budgets from different periods (see Figure 3). Given these new patches, we can define the pooled

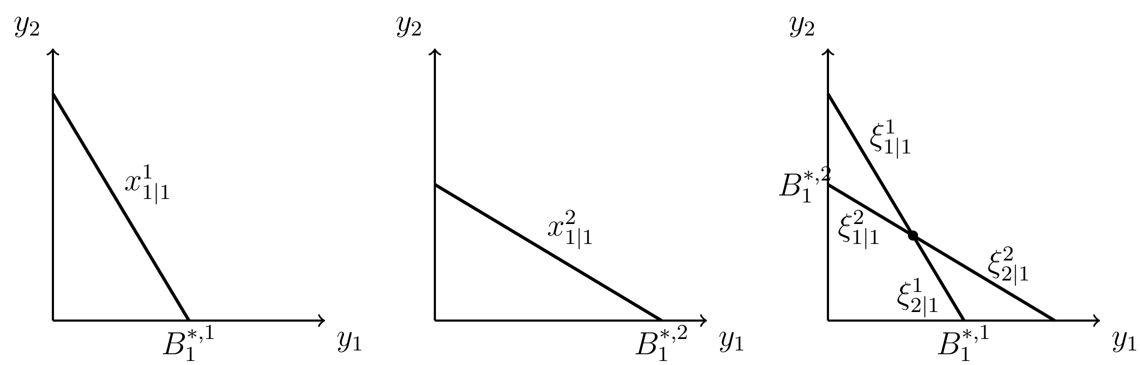

Figure 3 – $K = 2$ goods, $T = 2$ time periods, one budget per time period. The first and the second picture depict patches in 2 time periods. The third picture depicts new patches that arise after pooling the data.

stochastic function $\rho^{\text{pool}}\left(\xi_{k|j}^t\right)$ as the probability of observing someone picking from patch $\xi_{k|j}^t$. Next, we construct an example where ρ is rationalizable by DRUM but the corresponding

ρ^{pool} is not consistent with RUM (in the sense of Proposition 3). Consider the setting with $K = 2$ goods and $T = 2$ time periods. In each time period t, there is only one budget $B_1^{*,t}$. Assume that $B_1^{*,1} \neq B_1^{*,2}$ and $B_1^{*,1} \cup B_1^{*,2} \neq \emptyset$ (see Figure 3). Given that there is no budget variation for any given time period, there is only one choice path $\left(x_{1|1}^1, x_{1|1}^2\right)$. Thus, the trivial $\rho\left(\left(x_{1|1}^1, x_{1|1}^2\right)\right) = 1$ is rationalizable by DRUM. After pooling, since the budgets overlap, there are 4 patches (we assume that there is no intersection patch). Since there is only one choice path, DRUM does not impose any restrictions on the choice of individuals on these two budgets. As a result, we can take ν^1 and ν^2 from the DRUM definition such that $\rho^{\text{pool}}\left(\xi_{1|1}^2\right) + \rho^{\text{pool}}\left(\xi_{2|1}^1\right) > 1$. This ρ^{pool} cannot be consistent with RUM.

6. Counterfactuals in the Demand Setup

This section shows show how to conduct sharp counterfactual analyses within our framework in the demand setup.[28] The sharpness of our results follows from the fact that we have a full characterization of DRUM.[29] To simplify the exposition, we focus on the demand simple-setup (i.e., 2 intersecting budgets per period) for which we possess the $\mathcal{H}$-representation.

Given ρ in the time window $\mathcal{T}$, we want to bound some known function of counterfactual stochastic demands at the counterfactual time $T+1$. We assume that consumers face a pair of prices $p_{1,T+1}$ and $p_{2,T+1}$ that are known to the analyst at $T+1$. Let income in each period be 1. Denote the extended time window by $\mathcal{T}^{\mathrm{c}} = \mathcal{T} \cup \{T+1\}$. Similarly, the extended set of budget paths is denoted by $\mathbf{J}^{\mathrm{c}}$, and the extended vector representation of stochastic demand is denoted by ρ^{c}.

[28] Sharpness in this setting means that we can compute the tightest set of parameters that are consistent with the observed data and the model.

[29] See for early connections between nonparametric counterfactuals and specification testing Varian (1982, 1984), and Blundell, Browning and Crawford (2008), Norets and Tang (2014), Blundell, Kristensen and Matzkin (2014), Allen and Rehbeck (2019), Aguiar and Kashaev (2021), and Aguiar, Kashaev and Allen (2022) for recent examples in the analysis of demand, dynamic binary choice, and production.

Let $y^{\mathrm{c}}_{j_{T+1}}$ denote the counterfactual random demand of a consumer facing budget j_{T+1} at time $T+1$. That is,

$$y^{\mathrm{c}}_{j_{T+1}} = \arg\max_{y \in B^{T+1}_{j_{T+1}}} u^{T+1}(y),$$

where u^{T+1} is a random utility function at time $T+1$.

Definition 17 (Counterfactual marginal and conditional demands)**.** Given ρ, $x_{\mathbf{i}|\mathbf{j}}$, and budget $j_{T+1} \in \mathcal{J}^{T+1}$, the counterfactual conditional and marginal demands $\rho^{*}\left(\cdot|j_{T+1}, x_{\mathbf{i}|\mathbf{j}}\right)$ and $\rho^{**}\left(\cdot|j_{T+1}\right)$ are distributions over patches of j_{T+1} such that

$$\begin{aligned}
\rho^{*}_{j_{T+1}}\left(x^{T+1}_{i_{T+1}|j_{T+1}}|x_{\mathbf{i}|\mathbf{j}}\right) &= \rho^{\mathrm{c}}\left(\left(x^{t}_{i_t|j_t}\right)_{t\in\mathcal{T}^{\mathrm{c}}}\right) \Big/ \rho\left(x_{\mathbf{i}|\mathbf{j}}\right), \\
\rho^{**}_{j_{T+1}}\left(x^{T+1}_{i_{T+1}|j_{T+1}}\right) &= \sum_{x_{\mathbf{i}|\mathbf{j}}} \rho^{\mathrm{c}}\left(\left(x^{t}_{i_t|j_t}\right)_{t\in\mathcal{T}^{\mathrm{c}}}\right)
\end{aligned}$$

for any ρ^{c} that satisfies D-monotonicity, stability, and is such that

$$\rho\left(x_{\mathbf{i}|\mathbf{j}}\right) = \sum_{i_{T+1}\in\mathcal{I}^{T+1}_{j_{T+1}}} \rho^{\mathrm{c}}\left(\left(x^{t}_{i_t|j_t}\right)_{t\in\mathcal{T}^{\mathrm{c}}}\right).$$

The counterfactual conditional and marginal distributions fully characterize the choices of consumers in counterfactual situations, thus allowing us to compute sharp bounds for the expectation of any function of y^{c}. For a given measurable function $g : X \to \mathbb{R}$, let

$$\begin{aligned}
\underline{g}\left(x^{t}_{i_t|j_t}\right) &= \inf_{y\in x^{t}_{i_t|j_t}} g(y), \\
\overline{g}\left(x^{t}_{i_t|j_t}\right) &= \sup_{y\in x^{t}_{i_t|j_t}} g(y).
\end{aligned}$$

be the smallest and the largest value g can take over the patch $x^{t}_{i_t|j_t}$.

Proposition 4. Given ρ, $x_{\mathbf{i}|\mathbf{j}}$, and budget $j_{T+1} \in \mathcal{J}^{T+1}$,

$$\inf_{\rho^{*}_{j_{T+1}}} \sum_{i\in\mathcal{I}^{T+1}_{j_{T+1}}} \rho^{*}_{j_{T+1}}\left(x^{T+1}_{i|j_{T+1}}|x_{\mathbf{i}|\mathbf{j}}\right)\underline{g}\left(x^{T+1}_{i|j_{T+1}}\right) \le \mathbb{E}\left[g\left(y^{\mathrm{c}}_{j_{T+1}}\right)|x_{\mathbf{i}|\mathbf{j}}\right] \le \sup_{\rho^{*}_{j_{T+1}}} \sum_{i\in\mathcal{I}^{T+1}_{j_{T+1}}} \rho^{*}_{j_{T+1}}\left(x^{T+1}_{i|j_{T+1}}|x_{\mathbf{i}|\mathbf{j}}\right)\overline{g}\left(x^{T+1}_{i|j_{T+1}}\right),$$

$$\inf_{\rho^{**}_{j_{T+1}}} \sum_{i\in\mathcal{I}^{T+1}_{j_{T+1}}} \rho^{**}_{j_{T+1}}\left(x^{T+1}_{i|j_{T+1}}\right)\underline{g}\left(x^{T+1}_{i|j_{T+1}}\right) \leq \mathbb{E}\left[g\left(y^{c}_{j_{T+1}}\right)\right] \leq \sup_{\rho^{**}_{j_{T+1}}} \sum_{i\in\mathcal{I}^{T+1}_{j_{T+1}}} \rho^{**}_{j_{T+1}}\left(x^{T+1}_{i|j_{T+1}}\right)\overline{g}\left(x^{T+1}_{i|j_{T+1}}\right),$$

where the infimum and supremum are taken over all possible counterfactual marginal and conditional distributions.

Note that our results are complementary to those of Kitamura and Stoye (2019) that predict counterfactual stochastic demand for a new budget in a given cross-section using static RUM. We can use their techniques here as well. This section instead focuses on the counterfactual prediction in the time-dimension allowing dynamic preference change.

7. Empirical Application: Binary Menus of Lotteries

We study a sample of experimental subjects studied in ABKK that was surveyed in the MTurk platform between August 25, 2018 and September 17, 2018. We find evidence that deterministic rationality fails to explain the totality of the sample behavior, yet consistency with DRUM cannot be rejected. Moreover, restricting the linear orders in DRUM to those consistent with Expected Utility (DRUM-EU, Frick *et al.*, 2019) cannot explain the population behavior.

The sample contains $N = 2135$ decision makers (DMs). The grand choice set is the same across time $X^t = \{l_1, l_2, l_3\}$. These lotteries are defined over the set of prizes $Z = \{0, 10, 30, 50\}$ in tokens. The lotteries are $l_1 = (1/2, 0, 0, 0, 1/2)$, $l_2 = (0, 1/2, 1/2, 0)$ and $l_3 = \frac{1}{2}l_1 + \frac{1}{2}l_2$. The binary menus are $\{l_1, l_2\}, \{l_1, l_3\}, \{l_2, l_3\}$ at every $t \in \mathcal{T}$. In the experiment described in ABBK, consumers face one of these three menus uniformly at random and in uniformly random order. No DM faces the same binary menu twice. That means that there are 6 menu paths. Note that the design ensures that the probability of facing any of the 6 menu paths is uniform across DMs. Payments are made at random for one of the choices made by the DMs.

(For details of the payment, recruitment and sample demographics see ABBK.)

Concerns about limited consideration/attention are not first-order here because these binary menus were shown to DMs under a low-cost-of-attention treatment (see ABKK for details). In addition, DMs already faced a task focused on varying the attention cost as described in ABKK before facing the binary menus task. This means DMs are familiar with the lotteries when they face the binary menus. In addition, we do not have concerns about measurement error due to the experimental design and discrete choice nature of the data. This means we can focus on testing deterministic rationality at the individual level in this panel of choices versus DRUM without confounding due to inattention/limited consideration and measurement error.

Testing Deterministic Rationality

We test deterministic rationality at the individual DM level. In particular, we check SARP. SARP implies that we cannot have that $l_i \succ^* l_i'$, $l_i' \succ^* l_i''$ and $l_i'' \succ^* l_i$ with $l_i \succ^* l_i'$ defined as l_i is picked out of $\{l_i, l_i'\}$). We observe that 92% of DMs are consistent with SARP in our sample. Given that the power of this experiment to detect violations of SARP is low, we consider a 8% rejection rate to represent a significant fraction of DMs. Formally, the whole sample of DMs is not consistent with (static) utility maximization.

Testing DRUM

We compute the sample analogue of ρ, $\hat{\rho}$, following the methodology described in KS and ABKK. We implement the statistical test described in KS and ABKK to test the null that $\rho = A\nu$ for $\nu \geq 0$. A is computed in two steps. First, we compute the matrix A^t whose columns correspond to all possible linear orders on X^t and whose rows correspond to all possible lotteries from each menu. Observe that A^t is a square matrix of dimension 6 for all

$t \in \mathcal{T}$ such that $A_T = \otimes_{t=1}^{3} A^t$. In this application, we do not observe all possible choice paths since DMs never see repeated binary menus by design. Thus, A is obtained by extracting the submatrix of A_T that corresponds to observed menu paths (6 menu paths out of 27). We cannot reject the null hypothesis of consistency with DRUM (p-value is 0.72). Our finding confirms that a sample can be consistent with DRUM even when a significant fraction of DMs are not consistent with static utility maximization. We believe our results are compatible with the findings of Kurtz-David *et al.* (2019) that document that DMs fail deterministic rationality (in a different choice domain) because they make mistakes in evaluating the utility of lotteries. These mistakes can be interpreted as a form of dynamic random taste shocks. Indeed, DMs may get fatigued or learn, so their mistakes are correlated in time but with draws from the same distribution across choice paths and hence consistent with DRUM. The alleviate the concerns about the finite sample power of our test, we perform Monte Carlo experiments mimicking the setup of this application and find evidence that our test has high power even in small samples.

Testing DRUM-EU

Kashaev and Aguiar (2022b) propose a methodology to test for the null hypothesis of consistency with the Expected Utility version of RUM (RUM-EU). Their approach restricts the set of preference orders in the static test of RUM (i.e. preference orders consistent with static utility maximization, to those that are further consistent with expected utility maximization). We can apply the same idea to our dynamic setting to test the null hypothesis of consistency with DRUM-EU. That is, we define DRUM-EU as DRUM with the additional restriction that the set of utilities is consistent with Expected Utility. We reject the null hypothesis that ρ is consistent with DRUM-EU (p-value is less than 0.001). We note that this restriction on preferences is the focus of the work of Frick *et al.* (2019).

Testing Procedure

We set the number of bootstrap samples to $b = 999$. $N_{\mathbf{j}}$ denote the number of DMs on menu path $\mathbf{j} \in \mathbf{J}$. By construction, we have $\sum_{\mathbf{j}\in\mathbf{J}} N_{\mathbf{j}} = 2135$. For each bootstrap sample b^*_s, $s = 1, \ldots, b$, and each menu path $\mathbf{j} \in \mathbf{J}$, we draw $N_{\mathbf{j}}$ DMs with replacement from the observed dynamic stochastic demand $\hat{\rho}_{\mathbf{j}}$ to obtain the dynamic stochastic bootstrap demand $\hat{\rho}^*_{\mathbf{j},s}$. Following KS, the tuning parameter is set to $\tau_N = \sqrt{\log(\underline{N}_{\mathbf{j}})/\underline{N}_{\mathbf{j}}}$, where $\underline{N}_{\mathbf{j}}$ represents the smallest sample size across all menu paths. The test statistic $T^*_{N,s}$ of each bootstrap sample b^*_s is obtained following the statistical procedure in KS. We can obtain the p-value and critical values using the bootstrapped distribution of the test statistics.

8. Monte Carlo Simulations: Statistical Test of DRUM

Here we provide a Monte Carlo study to evaluate the performance of KS's test when applied to DRUM in finite samples. We consider both the demand setup and the binary menus setup.

8.1. Power Analysis: Demand Setup

We consider the demand simple-setup with $K = T = J^t = 2$. We set the number of DMs per choice path to $N_{\mathbf{i}|\mathbf{j}} \in \{50, 500, 5000\}$ and the number of simulations for each data generating process (DGP) to $R = 1000$. The critical value for each test statistic is computed using $b = 999$ bootstrap samples. As recommended in KS, the tuning parameter τ_N is set to $\tau_N = \sqrt{\log(4N_{\mathbf{i}|\mathbf{j}})/4N_{\mathbf{i}|\mathbf{j}}}$ (given that there are 4 choice paths in every budget path, $4N_{\mathbf{i}|\mathbf{j}}$ is the sample size of each budget path).

First, we consider a dynamic random Cobb-Douglas utility model. The utility function is given by

$$u_t(y_t) = y_{1,t}^{\alpha_t} y_{2,t}^{1-\alpha_t},$$

where $\alpha_t \in (0,1)$. Budgets in both periods are the same and correspond to prices $(2,1)'$ and $(1,2)'$ with an expenditure of 1. We consider 2 DGPs for random $\alpha = (\alpha_1, \alpha_2)'$.

$$\text{DGP1: } \alpha_1 \sim U[0,1]; \quad \alpha_2 = \max\{\min\{0.9\alpha_1 + \epsilon_1, 1\}, 0\},\ \epsilon_1 \sim N(0,25)$$

$$\text{DGP2: } \alpha_t = \arctan(\varepsilon_t)/\pi + 1/2,\ t = 1,2; \quad \varepsilon = (\varepsilon_1, \varepsilon_2)' \sim N(0,V)$$

where

$$V = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

Both DGPs are consistent with DRUM. The rejection rates at the 5% significance level for all 3 sample sizes and both DGPs are presented in Table 11. The rejection rates are close to 5%

DGP	$N_{i\|j}$	Rejection rate, %
DGP1	50	3.4
	500	4.3
	5000	5.1
DGP2	50	3.7
	500	4.6
	5000	5.4

Table 11 – Every entry represents the rejection rate at the 5% significance level and is computed from 1000 simulations and 999 bootstraps per simulation.

even for small sample sizes. To analyze the finite sample power of the test, we consider the DGP from Table 10. Recall that this ρ fails both D-monotonicity and stability. The rejection rate is 100% for all sample sizes. It is remarkable that ρ in Table 10 has marginal probabilities consistent with RUM. Yet, even at small sample sizes such as $N_{\mathbf{i}|\mathbf{j}} = 50$ the rejection rate is 100%. These simulations show that KS's test for DRUM has good size and power properties in finite samples in the demand setup.

8.2. Power Analysis: Mimicking the Empirical Application

We provide a Monte Carlo study to evaluate the performance of KS's test in a simulated environment mimicking our application with binary menus. This exercise is important because the number of observations per budget path is moderate. Hence, the asymptotic performance of the statistical test derived in the previous section may not translate to our application. We consider 3 data generating processes given by

$$\rho_1^t = \begin{bmatrix} 1/5 \\ 4/5 \\ 4/5 \\ 1/5 \\ 1/5 \\ 4/5 \end{bmatrix}, \ \rho_2^t = \begin{bmatrix} 1/5 \\ 4/5 \\ 1/2 \\ 1/2 \\ 1/5 \\ 4/5 \end{bmatrix}, \ \rho_3^t = \begin{bmatrix} 1/4 \\ 3/4 \\ 2/4 \\ 2/4 \\ 1/4 \\ 3/4 \end{bmatrix}.$$

Recall that the $\mathcal{H}$-representation of RUM is given by $H^t\rho \geq 0$, where H^t is given by Table 2 in our application. It is easy to check that

$$H^t\rho_1^t = [-0.4, 1.4, 1.4, -0.4, -0.4, 1.4]'$$

$$H^t\rho_2^t = [-0.1, 1.1, 1.1, -0.1, -0.1, 1.1]'$$

$$H^t\rho_3^t = [0, 1, 1, 0, 0, 1]'.$$

The dynamic extension of H^t is obtained from the Kronecker product of H^t, $H = \otimes_{t\in\mathcal{T}} H^t$. Likewise, the dynamic version of ρ_i^t is obtained from the Kronecker product of ρ_i^t, $\rho_i = \otimes_{t\in\mathcal{T}}\rho_i^t$, $i \in \{1, 2, 3\}$. Note that the first two DGPs are inconsistent with RUM while the third DGP is consistent with RUM. In the same way, the dynamic version of the first two DGPs are inconsistent with DRUM while the third one is consistent with DRUM. Specifically, observe that the size of the violations of DRUM are larger in the first DGP than in the second DGP and that the third DGP is a knife-edge case.

For the current analysis, we consider the same setup as in our application with $K = T = J^t = 3$. We set the number of consumers per budget path to $N_{\mathbf{j}} \in \{10, 175, 350\}$. This choice aims to be representative of the number of consumers per budget path in our application and to be informative about the small sample performance of the statistical test. We set the number of simulations for each data generating process (DGP) to $R = 1000$. The critical value for each test statistic is computed using $b = 999$ bootstrap samples. As recommended by KS, the tuning parameter τ_N is set to $\tau_N = \sqrt{\log(N_{\mathbf{j}})/N_{\mathbf{j}}}$.

The results are obtained using the test of KS based on the $\mathcal{V}$-representation of the model. The rejection rates at the 5% significance level for all 3 samples sizes and each DGP are presented in Table 12. As expected, false positives are less likely under the first DGP than the second DGP. Also, the third DGP shows that false negatives quickly attain the desired target level as the sample size grows. Overall, the results of Table 12 show that the statistical test performs very well even in small samples. In that sense, the nonrejection of DRUM in our application is unlikely to be the byproduct of a lack of power.

DGP	N_j	Rejection rate, %
DGP1	10	100
	175	100
	350	100
DGP2	10	25.3
	175	99.5
	350	100
DGP3	10	13.6
	175	6.0
	350	5.3

Table 12 – Every entry represents the rejection rate at the 5% significance level and is computed from 1000 simulations and 999 bootstraps per simulation.

9. Conclusion

We have fully characterized DRUM, a new model of consumer behavior when we observe a panel of choices from budget paths. In contrast to the static utility maximization framework, DRUM does not require the assumption that DMs or consumers keep their preferences stable over time. This generality is essential because the static utility maximization framework often fails to explain the behavior of individuals.

Our characterization works for any finite collection of choice paths in any finite time window. The characterization can be applied directly to existing panel consumption datasets using the statistical tools in KS. Moreover, our simple-setup characterization showcases that DRUM implies a richer set of behavioral restrictions on the panel of choices than RUM, alleviating some concerns about the empirical bite of the latter in a richer domain. These features position DRUM in-between Samuelson-Afriat's and McFadden-Richter's frameworks combining their strengths and reducing their weaknesses.

We have introduced to economics a generalization of the Weyl-Minkowski theorem for cones. This result is the basis of a recursive characterization of DRUM in the demand and abstract choice setups. This new mathematical result will be helpful beyond DRUM to obtain analogous generalizations of bounded rational models of stochastic choice.

References

Adams, A., Blundell, R., Browning, M. and Crawford, I. (2015). *Prices versus preferences: taste change and revealed preference*. Tech. rep., IFS Working Papers. 1, 2

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *Interna-*

tional economic review, **8** (1), 67–77. (document), 1

— (1973). On a system of inequalities in demand analysis: an extension of the classical method. *International economic review*, pp. 460–472. 5.1

Aguiar, V. H., Boccardi, M. J., Kashaev, N. and Kim, J. (2023). Random utility and limited consideration. *Quantitative Economics*, **14** (1), 71–116. 1

— and Kashaev, N. (2021). Stochastic revealed preferences with measurement error. *The Review of Economic Studies*, **88** (4), 2042–2093. 1, 2, 29

—, — and Allen, R. (2022). Prices, profits, proxies, and production. *Journal of Econometrics*. 29

— and Serrano, R. (2021). Cardinal revealed preference: Disentangling transitivity and consistent binary choice. *Journal of Mathematical Economics*, **94**, 102462. 1

Aguirregabiria, V. and Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, **156** (1), 38–67. 1

Ahn, D., Choi, S., Gale, D. and Kariv, S. (2014). Estimating ambiguity aversion in a portfolio choice experiment. *Quantitative Economics*, **5** (2), 195–223. 1

Akesaka, M., Eibich, P., Hanaoka, C. and Shigeoka, H. (2021). Temporal instability of risk preference among the poor: Evidence from payday cycles. *American Economic Journal: Applied Economics*, forthcoming. 1

Allen, R. and Rehbeck, J. (2019). Identification with additively separable heterogeneity. *Econometrica*, **87** (3), 1021–1054. 29

Apesteguia, J., Ballester, M. A. and Gutierrez-Daza, A. (2022). Random discounted expected utility. 1

Aubrun, G., Lami, L., Palazuelos, C. and Plávala, M. (2021). Entangleability of cones. *Geometric and Functional Analysis*, **31** (2), 181–205. 1, 4, 4, 18, 10.4

—, Müller-Hermes, A. and Plávala, M. (2022). Monogamy of entanglement between cones. *arXiv preprint arXiv:2206.11805*. 1, 4, 15, 4, 18, 10.4

Bandyopadhyay, T., Dasgupta, I. and Pattanaik, P. K. (1999). Stochastic revealed preference and the theory of demand. *Journal of Economic Theory*, **84** (1), 95–110. 1, 4.1

—, — and — (2004). A general revealed preference theorem for stochastic demand behavior. *Economic Theory*, **23** (3), 589–599. 4.1

Becker, G. S. and Murphy, K. M. (1988). A theory of rational addiction. *Journal of political Economy*, **96** (4), 675–700. 1

Block, H. and Marschak, J. (1960). Random orderings and stochastic theories of responses". in i. olkin, s. ghurye, w. hoeffding, w. madow, and h. man (eds) contributions to probability and statistics, stanford university press. 1, 4

Blundell, R., Browning, M. and Crawford, I. (2008). Best nonparametric bounds on demand responses. *Econometrica*, **76** (6), 1227–1262. 29

—, Kristensen, D. and Matzkin, R. (2014). Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics*, **179** (2), 112–127. 29

Border, K. (2007). Introductory notes on stochastic rationality. *California Institute of Technology*. 3, 10.2

Brocas, I., Carrillo, J. D., Combs, T. D. and Kodaverdian, N. (2019). Consistency in simple vs. complex choices by younger and older adults. *Journal of Economic Behavior & Organization*, **157**, 580–601. 1

Browning, M. (1989). A nonparametric test of the life-cycle rational expections hypothesis. *International Economic Review*, pp. 979–992. 1, 2.4

Cattaneo, M. D., Ma, X., Masatlioglu, Y. and Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, **128** (7), 2796–2836. 4

Chambers, C. P., Masatlioglu, Y. and Turansick, C. (2021). Correlated choice. *arXiv preprint arXiv:2103.05084*. 1, 2.1, 2.2, 4, 4, 21, 4.2

Cherchye, L., Demuynck, T., De Rock, B. and Vermeulen, F. (2017). Household consumption when the marriage is stable. *American Economic Review*, **107** (6), 1507–1534. 1

Choi, S., Fisman, R., Gale, D. M. and Kariv, S. (2007). Revealing preferences graphically: an old method gets a new tool kit. *American Economic Review*, **97** (2), 153–158. 1

—, Kariv, S., Müller, W. and Silverman, D. (2014). Who is (more) rational? *The American Economic Review*, **104** (6), 1518–1550. 1

Dasgupta, I. and Pattanaik, P. K. (2007). 'regular'choice and the weak axiom of stochastic revealed preference. *Economic Theory*, **31** (1), 35–50. 4.1

de Bruyn, J. v. D. (2020). Tensor products of convex cones, part ii: Closed cones in finite-dimensional spaces. *arXiv preprint arXiv:2009.11843*. 4

Dean, M. and Martin, D. (2016). Measuring rationality with the minimum cost of revealed preference violations. *Review of Economics and Statistics*, **98** (3), 524–534. 1

Deb, R., Kitamura, Y., Quah, J. K.-H. and Stoye, J. (2021). Revealed price preference: Theory and statistical analysis. *Review of Economic Studies*. 5, 6, 1, 2, 2.4, 2.4, 2.4, 5.3

Demuynck, T. and Verriest, E. (2013). Iâ€™ll never forget my first cigarette: a revealed preference analysis of the "habits as durables" model. *International Economic Review*, **54** (2), 717–738. 1

Dogan, S. and Yildiz, K. (2022). Every choice function is pro-con rationalizable. *Operations Research*. 4, 10.5

Doherty, A. C., Parrilo, P. A. and Spedalieri, F. M. (2004). Complete family of separability criteria. *Physical Review A*, **69** (2), 022308. 4

Dridi, T. (1980). Sur les distributions binaires associes des distributions ordinales. *Mathmatiques et Sciences Humaines*, **69**, 15–31. 9

Echenique, F., Lee, S. and Shum, M. (2011). The money pump as a measure of revealed preference violations. *Journal of Political Economy*, **119** (6), 1201–1223. 1

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology*, **18** (1), 52–72. 4, 10.7

Fang, Z., Santos, A., Shaikh, A. M. and Torgovitsky, A. (2023). Inference for large-scale linear systems with known coefficients. *Econometrica*, **91** (1), 299–327. 26

Fishburn, P. C. (1998). Stochastic utility. In S. Barberá, P. J. Hammond and C. Seidl (eds.), *Handbook of Utility Theory*, pp. 272–311. 4.2

Frick, M., Iijima, R. and Strzalecki, T. (2019). Dynamic random utility. *Econometrica*, **87** (6), 1941–2002. 1, 7

Fudenberg, D. and Strzalecki, T. (2015). Dynamic logit with choice aversion. *Econometrica*, **83** (2), 651–691. 1

Gauthier, C. (2018). Nonparametric identification of discount factors under partial efficiency. *Working paper*. 1, 2

— (2021). Price search and consumption inequality: Robust, credible, and valid inference. *Working paper*. 2

Guiso, L., Sapienza, P. and Zingales, L. (2018). Time varying risk aversion. *Journal of Financial Economics*, **128** (3), 403–421. 1

Halevy, Y. and Mayraz, G. (2022). Identifying rule-based rationality. *Review of Economics and Statistics*, pp. 1–44. 1

Hoderlein, S. and Stoye, J. (2014). Revealed preferences in a heterogeneous population. *Review of Economics and Statistics*, **96** (2), 197–213. 1, 4

Im, C. and Rehbeck, J. (2021). Non-rationalizable individuals, stochastic rationalizability, and sampling. *Available at SSRN 3767994*. 1, 20

Kashaev, N. and Aguiar, V. H. (2022a). A random attention and utility model. *Journal of Economic Theory*, **204**, 105487. 4

— and — (2022b). Random rank-dependent expected utility. *Games*, **13** (1), 13. 7, 7

Kawaguchi, K. (2017). Testing rationality without restricting heterogeneity. *Journal of Econometrics*, **197** (1), 153–171. 1, 2.3, 2.3, 3, 10.2

Kitamura, Y. and Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, **86** (6), 1883–1909. 1, 6, 2.3, 4, 4, 5.3

— and — (2019). Nonparametric counterfactuals in random utility models. *arXiv preprint arXiv:1902.08350*. 6

Kurtz-David, V., Persitz, D., Webb, R. and Levy, D. J. (2019). The neural computation of inconsistent choice behavior. *Nature communications*, **10** (1), 1583. 1, 7

Li, R. (2021). An axiomatization of stochastic utility. *arXiv preprint arXiv:2102.00143*. 1, 2.1, 2.2, 4, 21, 4.2

Lu, J. and Saito, K. (2018). Random intertemporal choice. *Journal of Economic Theory*, **177**, 780–815. 1

McCausland, W. J., Davis-Stober, C., Marley, A. A., Park, S. and Brown, N. (2020). Testing the random utility hypothesis directly. *The Economic Journal*, **130** (625), 183–207. 2, 2.3

McFadden, D. and Richter, M. K. (1990). Stochastic rationality and revealed stochastic preference. *Preferences, Uncertainty, and Optimality, Essays in Honor of Leo Hurwicz, Westview Press: Boulder, CO*, pp. 161–186. (document), 1, 3, 5, 10.10

McFadden, D. L. (2005). Revealed stochastic preference: a synthesis. *Economic Theory*, **26** (2), 245–264. 3, 5

Norets, A. and Tang, X. (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies*, **81** (3), 1229–1262. 29

Porter, M. and Adams, A. (2016). For love or reward? characterising preferences for giving to parents in an experimental setting. *The Economic Journal*, **126** (598), 2424–2445. 2, 2.3

Ray, D. and Robson, A. (2018). Certified random: A new order for coauthorship. *American Economic Review*, **108** (2), 489–520. ∗

Saito, K. (2017). Axiomatizations of the mixed logit model. 4, 10.5

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, **5** (17), 61–71. 1, 5.1

Smeulders, B., Cherchye, L. and De Rock, B. (2021). Nonparametric analysis of random utility models: computational tools for statistical testing. *Econometrica*, **89** (1), 437–455. 3, 3, 4

Stoye, J. (2019). Revealed stochastic preference: A one-paragraph proof and generalization. *Economics Letters*, **177**, 66–68. 4

Strzalecki, T. (2021). *Stochastic Choice*. Mimeo. 1, 4

Turansick, C. (2022). Identification in the random utility model. *Journal of Economic Theory*, **203**, 105489. 4.2, 25, 4.2

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica: Journal of the Econometric Society*, pp. 945–973. 29

— (1984). The nonparametric approach to production analysis. *Econometrica: Journal of the Econometric Society*, pp. 579–597. 29

10. Proofs

10.1. Proof of Lemma 1

We adapt the proof of Theorem 3.1 in KS for RUM for the dynamic case. Our proof uses profiles of nonstochastic demand. For each time period $t \in \mathcal{T}$, we define nonstochastic demand types as in KS: $(\theta_1^t, \cdots, \theta_{J^t}^t) \in B_1^t \times \cdots \times B_{J^t}^t$. This system of types is rationalizable if $\theta_j^t \in \arg\max_{y \in B_j^t} u^t(y)$ for $j = 1, \cdots, J^t$ for some utility function u^t.

Then, we form any given nonstochastic demand profile by stacking up the demand types in a budget path $\mathbf{j}$ as $\theta_{\mathbf{j}} = (\theta_{j_t}^t)_{j_t \in \mathbf{j}}$.

Fix ρ. For a fixed $t \in \mathcal{T}$, let the set $\mathcal{Y}_t^{**}$ collect the geometric center point of each patch. Let ρ^{**} be the unique dynamic stochastic demand system concentrated on $\mathcal{Y}_t^{**}$ for all $t \in \mathcal{T}$. KS established that demand systems can be arbitrarily perturbed within patches in a given time period t such that ρ is rationalizable by DRUM if and only if ρ^{**} is. It follows that the rationalizability of ρ can be decided by checking whether there exists a mixture of nonstochastic demand profiles supported on $\mathcal{Y}_t^{**}$ for all $t \in \mathcal{T}$.

Since we have assumed a finite number of budgets and time periods, there will be a finite number of budget paths. That is, using our notation, we have $|\mathbf{J}|$ budget paths. Also, because $\mathcal{Y}_t^{**}$ is finite for all $t \in \mathcal{T}$, there are finitely many nonstochastic demand profiles. Noting that these demand profiles are characterized by binary vector representations corresponding to columns of A_T, the statement of the theorem follows immediately.

10.2. Proof of Theorem 1

To prove $(i) \iff (ii) \iff (iii)$, we adapt the proof of Theorem 3.1 in KS for RUM for the dynamic case. Our proof uses nonstochastic linear order profiles. Then, up to this redefinition $(i) \iff (ii) \iff (iii)$ follows from their results.

The proof of $(i) \implies (iv)$ follows from Border (2007). The proof of $(iv) \implies (i)$ is analogous to the proof for the case of RUM in Border (2007) and Kawaguchi (2017). We just need to replace the system of equations in those proofs with the one we describe in Theorem 1 (ii). The rest of the proof follows from Farkas' lemma.

10.3. Proof of Proposition 1

For completeness we provide here the proof of Proposition 1. Let $L_T = \otimes_{t=1}^T L^t$ and $K_T = \otimes_{t=1}^T K^t$. Note that for any v, z and $\otimes_{t=1}^T K^t$ such that $\left(\otimes_{t=1}^T K_t\right) v = z$ is well-defined, we can construct V and Z such that columns of V and Z are subvectors[30] of v and z and

$$\left(\otimes_{t=1}^T K^t\right) v = z \iff K^T V \left(\otimes_{t=1}^{T-1} K^t\right)' = Z.$$

Recall that by definition, $L^t K^t v \geq 0$ for all $v \geq 0$. Hence,

$$\begin{aligned}
&\forall v \geq 0,\ L^1 K^1 v \geq 0 \implies \forall V \geq 0,\ L^2 K^2 V (L^1 K^1)' \geq 0 \iff \\
&\forall v \geq 0,\ (L^1 K^1 \otimes L^2 K^2) v \geq 0 \implies \forall V \geq 0,\ L^3 K^3 V (L^1 K^1 \otimes L^2 K^2)' \geq 0 \iff \\
&\forall v \geq 0,\ (\otimes_{t=1}^3 L^t K^t) v \geq 0 \implies \forall V \geq 0,\ L^4 K^4 V (\otimes_{t=1}^3 L^t K^t)' \geq 0 \implies
\end{aligned}$$

[30] A vector x is a subvector of $y = (y_j)_{j \in J}$, if $x = (y_j)_{j \in J'}$ for some $J' \subseteq J$.

$$\cdots \implies \forall v \geq 0,\ (\otimes_{t=1}^{T} L^t K^t)v \geq 0 \iff \forall v \geq 0,\ L_T K_T v \geq 0.$$

Hence,

$$\{K_T v \ :\ v \geq 0\} \subseteq \{z \ :\ L_T z \geq 0\}.$$

10.4. Proof of Theorem 2

The proof of the first statement follows directly from Theorem 1 in Aubrun *et al.* (2022). The proof of the moreover statement follows from Corollary 4 in Aubrun *et al.* (2021).

For self-containment we prove here directly the sufficiency in the moreover statement. Without loss of generality, assume that K^t has full column rank for all t except maybe $t = 1$. By assumptions of the theorem, K^t is proper and, thus, of full row rank for all t. Hence, K_T has full row rank and we can represent any z as a weighted sum of columns of K_T (some weights may be negative). That is, $\{v \ :\ K_T v = z\}$ is nonempty for any z. Take any z such that $L_T z \geq 0$. We want to show that there exists $v \geq 0$ such that $K_T v = z$. Towards a contradiction, assume that $v \not\geq 0$ for all $v \in \{v \ :\ K_T v = z\}$. Take any $v \in \{v \ :\ K_T v = z\}$. Then

$$L_T K_T v \geq 0 \implies L^T K^T V (L_{T-1} K_{T-1})' \geq 0,$$

where V is constructed the same way it was constructed in the proof of Proposition 1. Since K^T is invertible (full row and column rank), we have that $V(L_{T-1}K_{T-1})' \geq 0$. Take any row of V that has a negative component and call the transpose of this row v. Then

$$L_{T-1} K_{T-1} v \geq 0 \implies L^{T-1} K^{T-1} V (L_{T-2} K_{T-2})' \geq 0.$$

Hence, $L_{T-1}K_{T-1}V' \geq 0$. Repeating this step finitely many times we end up having $L^1 K^1 v \geq 0$ for some $v \not\geq 0$. Since K^1 may not have full column rank, we only can conclude that there exists $v^* \geq 0$ such that $K^1 v = K^1 v^*$. We can construct such $v^* \geq 0$ for all possible subvectors of the original v. Combine these v^*s into $\bar{v} \geq 0$. By definition, $K_T \bar{v} = z$ and $\bar{v} \geq 0$. The latter is not possible since it was assumed that $v \ngeq 0$ for all $v \in \{v : K_T v = z\}$. The contradiction completes the proof.

10.5. Proof of Theorem 3

First we show necessity of stability. By definition of DRUM, there exists a distribution over $\mathcal{U}$, μ, such that

$$\rho\left(\left(x_{i_t|j_t}\right)_{t\in\mathcal{T}}\right) = \int \prod_{t\in\mathcal{T}} \mathbb{1}\left(\arg\max_{y\in B^t_{j_t}} u^t(y) = x^t_{i_t|j_t}\right) d\mu(u)$$

for all $\mathbf{i}, \mathbf{j}$. Fix some $t' \in \mathcal{T}$, $x_{\mathbf{i}|\mathbf{j}}$, and $j_{t'} \in \mathcal{J}^{t'}$. Note that

$$\begin{aligned}
&\sum_{i\in\mathcal{I}^{t'}_{j_{t'}}} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right) = \\
&\sum_{i\in\mathcal{I}^{t'}_{j_{t'}}} \int \mathbb{1}\left(\arg\max_{y\in B^{t'}_{j_{t'}}} u^{t'}(y) = x^{t'}_{i|j_{t'}}\right) \prod_{t\in\mathcal{T}\setminus\{t'\}} \mathbb{1}\left(\arg\max_{y\in B^t_{j_t}} u^t(y) = x^t_{i_t|j_t}\right) d\mu(u) = \\
&\int \sum_{i\in\mathcal{I}^{t'}_{j_{t'}}} \mathbb{1}\left(\arg\max_{y\in B^{t'}_{j_{t'}}} u^{t'}(y) = x^{t'}_{i|j_{t'}}\right) \prod_{t\in\mathcal{T}\setminus\{t'\}} \mathbb{1}\left(\arg\max_{y\in B^t_{j_t}} u^t(y) = x^t_{i_t|j_t}\right) d\mu(u) = \\
&\int \prod_{t\in\mathcal{T}\setminus\{t'\}} \mathbb{1}\left(\arg\max_{y\in B^t_{j_t}} u^t(y) = x^t_{i_t|j_t}\right) d\mu(u),
\end{aligned}$$

where the last equality follows from $\arg\max_{y\in B^{t'}_{j_{t'}}} u^{t'}(y)$ being a singleton and $\{x^{t'}_{i|j_{t'}}\}_{i\in\mathcal{I}^{t'}_{j_{t'}}}$ being a partition. The right-hand side of the last expression does not depend on the choice of $j_{t'}$. Stability follows from t' and $x_{\mathbf{i}|\mathbf{j}}$ being arbitrary.

Next, we show that any stable ρ belongs to a linear span of columns of A_T. That is, the system $A_T v = \rho$ always has a solution and the cone generated by A_T is proper when restricted to stable ρ. Hence, Theorem 3 follows from Theorem 2.

Before we formally show the existence of a solution, let us first replicate the proof in the simple-setup with 2 time periods. Recall that

$$A^t = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

First, construct the matrix A^{t*} by removing the last row from A^t. That is, for every budget except the first one (the first two rows), remove the row that corresponds to the last patch of that budget (rows 3 and 4 correspond to the second budget). As a result,

$$A^{t*} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Put the removed row in the matrix A^{t-}. That is, $A^{t-} = (0\,1\,1)$. Note that $A^{t-} = G^t A^{t^*}$, where $G^t = (1\,1\,\,-1)$. Moreover, $A_T = A^1 \otimes A^2$ (see Table 3) can be partitioned into the matrix A^*_T that contains the rows generated by rows of A^{1*} and A^{2*} ($A^*_T = A^{1*} \otimes A^{2*}$), and the matrix A^-_T that contains the rest of the rows. That is,

$$A_T = \begin{pmatrix} A^{1*} \\ A^{1-} \end{pmatrix} \otimes \begin{pmatrix} A^{2*} \\ A^{2-} \end{pmatrix} = \begin{pmatrix} A^{1*} \otimes A^{2*} \\ A^{1*} \otimes A^{2-} \\ A^{1-} \otimes A^{2*} \\ A^{1-} \otimes A^{2-} \end{pmatrix} = \begin{pmatrix} A^*_T \\ A^-_T \end{pmatrix}.$$

Let ρ^* and ρ^- be the parts of ρ that correspond to rows of A^*_T and A^-_T. Since every element of ρ corresponds to some choice path, ρ^* does not contain choice paths that contain either

$x^1_{2|2}$ or $x^2_{2|2}$ (we removed one row from A^1 and one row from A^2). Similarly, ρ^- contains all choice paths where at least in one time period t a patch was removed from A^t.

Note that A^{t*}, $t \in \mathcal{T}$, has full row rank. Hence, $A^*_{\mathcal{T}}$, as a Kronecker product of full row rank matrices, is of full row rank as well. Thus, $v^* = A^{*\prime}_{\mathcal{T}}(A^*_{\mathcal{T}}A^{*\prime}_{\mathcal{T}})^{-1}\rho^*$ exists and solves $A^*_{\mathcal{T}}v = \rho^*$. If we show that $A^-_{\mathcal{T}}v^* = \rho^-$, then v^* solves $A_{\mathcal{T}}v = \rho$ as well. Note that,

$$A^{1*} \otimes A^{2-}v^* = \left(A^{1*} \otimes G^2 A^{2*}\right) v^* = \begin{pmatrix} G^2 & 0 & \dots \\ 0 & G^2 & \dots \\ \dots & \dots & \dots \\ \dots & 0 & G^2 \end{pmatrix} \left(A^{1*} \otimes A^{2*}\right) v^* = diag(G^2)\rho^*,$$

where $diag(L)$ is a block-diagonal matrix with matrix L being on the main diagonal. The vector ρ^* has 9 elements with the first 3 elements corresponding to choice paths that have $x^1_{1|1}$ and all possible patches that were not removed from $t = 2$. That is, the first 3 elements of ρ^* are $\rho\left(\left(x^1_{1|1}, x^2_{1|1}\right)\right)$, $\rho\left(\left(x^1_{1|1}, x^2_{2|1}\right)\right)$, and $\rho\left(\left(x^1_{1|1}, x^2_{1|2}\right)\right)$ (the patch $x^2_{2|2}$ was removed). Thus, the first element of $diag(G^2)\rho^*$ is

$$\rho\left(\left(x^1_{1|1}, x^2_{1|1}\right)\right) + \rho\left(\left(x^1_{1|1}, x^2_{2|1}\right)\right) - \rho\left(\left(x^1_{1|1}, x^2_{1|2}\right)\right) = \rho\left(\left(x^1_{1|1}, x^2_{2|2}\right)\right),$$

where the equality follows from stability of ρ. Similarly, the second element of $diag(G^2)\rho^*$ is

$$\rho\left(\left(x^1_{2|1}, x^2_{1|1}\right)\right) + \rho\left(\left(x^1_{2|1}, x^2_{2|1}\right)\right) - \rho\left(\left(x^1_{2|1}, x^2_{1|2}\right)\right) = \rho\left(\left(x^1_{2|1}, x^2_{2|2}\right)\right),$$

and the third element is $\rho\left(\left(x^1_{1|2}, x^2_{2|2}\right)\right)$. So v^* solves the equations with only $x^2_{2|2}$ dropped.

Next, consider $A^{1-} \otimes A^{2*}v^*$. Note that all objects we work with (e.g., $A_{\mathcal{T}}$ and $A^*_{\mathcal{T}}$) are defined as a function of $\mathcal{T}$. Hence, if we push the time period t to the very end (i.e., $1, \dots, t-1, t+1, \dots, T, t$), we still can define all objects for the new order of time labels. Let W^t (with inverse $W^{t,-1}$, which pushes the last element of $\mathcal{T}$ to t-th position) be a transformation that recomputes all objects for the time span where t is pushed to the end. For example, W^1 pushes the label $t = 1$ to the end of $\mathcal{T}$ (i.e., $\mathcal{T}$ becomes $\{2, 1\}$). Transformation W^t satisfies the following three properties: $W^t[C] = C$ if C does not depend on $\mathcal{T}$; $W^t[CD] = W^t[C]W^t[D]$ for any matrices C and D; and $W^t[\otimes_{t'\in\mathcal{T}} A^{t'*}] = \otimes_{t'\in\mathcal{T}\setminus\{t\}} A^{t'*} \otimes A^{t*}$. Hence,

$$
\begin{aligned}
&A^{1-} \otimes A^{2*} v^* = W^{1,-1}\left[W^1\left[\left(A^{1-} \otimes A^{2*}\right) v^*\right]\right] = W^{1,-1}\left[\left(A^{2*} \otimes A^{1-}\right) W^1\left[v^*\right]\right] = \\
&W^{1,-1}\left[diag(G^1)\left(A^{2*} \otimes A^{1*}\right) W^1\left[v^*\right]\right] = W^{1,-1}\left[diag(G^1) W^1\left[W^{1,-1}\left[\left(A^{2*} \otimes A^{1*}\right) W^1\left[v^*\right]\right]\right]\right] = \\
&W^{1,-1}\left[diag(G^1) W^1\left[\left(A^{1*} \otimes A^{2*}\right) v^*\right]\right] = W^{1,-1}\left[diag(G^1) W^1\left[\rho^*\right]\right].
\end{aligned}
$$

In words, $W^1[\rho^*]$ changes labels so that $t = 1$ is the last one and reshuffles elements of ρ^*, then $diag(G^1)W^1[\rho^*]$ computes probabilities of choice paths where $x^2_{2|2}$ were dropped. Finally, $W^{1,-1}$ returns the original labeling. So the result is the subvector of ρ^- where $x^1_{2,2}$ is dropped (relabelling changes $x^2_{2|2}$ to $x^1_{2|2}$). So v^* solves the equations where only $x^1_{2|2}$ is dropped.

Let Y^t be an operator such that $Y^t[\cdot] = W^{t,-1}[diag(G^t)W^t[\cdot]]$. That is, Y^t pushes t to the end, multiplies the resulting object by $diag(G^t)$ and then pushes label t back to its spot. Using operator Y^t we can deduce that

$$
\begin{aligned}
&A^{1-} \otimes A^{2-} v^* = Y^1\left[Y^2\left[\rho^*\right]\right] = \\
&\rho\left(\left(x^1_{2|2}, x^2_{1|1}\right)\right) + \rho\left(\left(x^1_{2|2}, x^2_{2|1}\right)\right) - \rho\left(\left(x^1_{2|2}, x^2_{1|2}\right)\right) = \rho\left(\left(x^1_{2|2}, x^2_{2|2}\right)\right),
\end{aligned}
$$

where the last equality follows from stability of ρ. Hence, the equation where both $x^1_{2|2}$ and $x^2_{2|2}$ were dropped is also solved by v^*.

Next, we generalize the above arguments for arbitrary T and A^t. Consider the following modification of A^t, $t \in \mathcal{T}$. From every menu, except the first one, we pick the last alternative and remove the corresponding row from A^t. Let A^{t*} denote the resulting matrix. Thus, matrix A^t can be partitioned into A^{t*} and A^{t-}, where rows of A^{t-} correspond to alternatives removed from A^t. Consider the first row of A^{t-}. It corresponds to the last alternative from the second menu at time t. Note that the sum of all rows that correspond to the same menu is equal to the row of ones. Hence, the first row of A^{t-} is equal to the sum of the rows that correspond to menu 1 minus the sum of the remaining rows in menu 2. That is, the first row of A^{t-} can be written as

$$(1,\ldots,1,-1,\ldots,-1,0,\ldots,0)A^{t*}.$$

Similarly, the second row of A^{t-} can be written as

$$(1,\ldots,1,0,\ldots,0,-1,\ldots,-1,0,\ldots,0)A^{t*}.$$

In matrix notation, we can rewrite A^{t-} as $A^{t-}=G^tA^{t*}$, where G^t is the matrix with the k-th row having the elements that correspond to the alternatives from the first menu at time t are equal to 1, the elements that correspond to the alternatives from the k-th menu are equal to -1, and the rest of elements are equal to 0.

Next note that, up to a permutation of rows, A_T can be partitioned into $A^*_T=\otimes_{t\in\mathcal{T}}A^{t*}$ and matrices of the form $\otimes_{t\in\mathcal{T}}C^t$, where $C^t\in\{A^{t*},A^{t-}\}$, with $C^t=A^{t-}$ for at least one t. We will stack all these matrices into A^-_T. Next, let ρ^* denote the subvector of ρ that corresponds to choice paths that do not contain any of the alternatives removed from A^t, $t\in\mathcal{T}$. Thus, $\rho=(\rho^{*\prime},\rho^{-\prime})'$, where ρ^- corresponds to all elements of ρ that contain at least one of the removed alternatives. As a result, we can split the original system into two: $A^*_Tv=\rho^*$ and $A^-_Tv=\rho^-$.

Consider the system $A^*_Tv=\rho^*$. We formally prove later that A^{t*} has full row rank for all t. Then A^*_T is also of full row rank and, hence, $A^*_TA^{*\prime}$ is invertible and $v^*=A^{*\prime}\left(A^*_TA^{*\prime}\right)^{-1}\rho^*$ solves the system. If we show that

$$A^-_Tv^*=\rho^-,$$

then we prove that $A_Tv=\rho$ always has a solution, which will complete the proof.

Note that A^-_T consists of the blocks of the form $\otimes_{t\in\mathcal{T}}C^t$, where $C^t\in\{A^{t*},A^{t-}\}$ and $C^t=A^{t-}$

for at least one t. Next note that for any A, B, and C

$$A \otimes (BC) = diag(B)(A \otimes C),$$

where $diag(B)$ is the block-diagonal matrix constructed from B. Indeed,

$$A \otimes (BC) = \begin{pmatrix} A_{11}BC & A_{12}BC & \dots \\ A_{21}BC & A_{22}BC & \dots \\ \dots & \dots & \dots \end{pmatrix} = \begin{pmatrix} B & 0 & \dots \\ 0 & B & \dots \\ \dots & \dots & B \end{pmatrix} \begin{pmatrix} A_{11}C & A_{12}C & \dots \\ A_{21}C & A_{22}C & \dots \\ \dots & \dots & \dots \end{pmatrix}$$
$$= diag(B)(A \otimes C)$$

First, consider $\otimes_{t\in\mathcal{T}} C^t$, where $C^t \in \{A^{t*}, A^{t-}\}$ and $C^t = A^{t-}$ for only one t. Hence,

$$\otimes_{t'\in\mathcal{T}} C^{t'} v^* = W^{t,-1}\left[diag(G^t)W^t\left[\rho^*\right]\right] = Y^t[\rho^*].$$

Note that because ρ is stable, $diag(G^T)\rho^*$ is the subvector of ρ^- that corresponds to choice paths that contain one of the removed alternatives from the last period only. So, $W^t\left[\rho^*\right]$ first pushes the period t to the very end, then $diag(G^t)W^t\left[\rho^*\right]$ computes the elements of ρ^-, and finally $W^{t,-1}\left[diag(G^t)W^t\left[\rho^*\right]\right]$ moves the time period t back to its place.

Next, consider $\otimes_{t\in\mathcal{T}} C^t$, where $C^t \in \{A^{t*}, A^{t-}\}$ and $C^t = A^{t-}$ and $C^{t'} = A^{t'-}$ for two distinct t, t'. Similarly to the previous case,

$$\otimes_{t'\in\mathcal{T}} C^{t'} v^* = W^{t,-1}\left[diag(G^t)W^t\left[W^{t',-1}\left[diag(G^{t'})W^{t'}\left[\rho^*\right]\right]\right]\right] = Y^t[Y^{t'}[\rho^*]] = Y^t \circ Y^{t'}[\rho^*],$$

where $Y^t \circ Y^{t'}$ denotes the composite operator. Again, $W^{t',-1}\left[diag(G^{t'})W^{t'}\left[\rho^*\right]\right]$ computes the subvector of ρ^- that corresponds to choice paths where an alternative from only one time t' was missing. Applying to the resulting vector $W^{t,-1}\left[diag(G^t)W^t\left[\cdot\right]\right]$ computes the subvector of ρ^- with alternatives missing from t and t' only. Repeating the arguments for all possible rows of A_T^-, we obtain that

$$\otimes_{t'\in\mathcal{T}} C^{t'} v^* = \circ_{t':C^{t'}=A^{t'-}} Y^{t'}[\rho^*]$$

and, thus, $A_T^- v^* = \rho^-$. Hence, v^* is a solution to $A_T v = \rho$.

It is left to show that A_t^* is a full row rank matrix for all t. To do so, we first prove the same result for a more general version of static RUM with "virtual" budgets introduced in Section 4.2.

Let $\bar{\mathcal{R}}^t$ be the set of all linear orders on $\mathbf{X}^t$. For any $j_t \in \bar{\mathcal{J}}^t, i_t \in \mathcal{I}_{j_t}^t$, and $\succ \in \bar{\mathcal{R}}^t$ let

$$\bar{a}_{\succ}^t = \left(\mathbb{1}\left(x_{i_t|j_t}^t \succ x,\ \forall x \in B_{j_t}^t\right)\right)_{j_t\in\bar{\mathcal{J}}^t, i_t\in\mathcal{I}_{j_t}^t}$$

be the vector of 0s and 1s that denote the best patch in every virtual budget. Analogously to A^t, let $\bar{A}^t$ denote the matrix which columns are $\{a_{\succ}^t\}_{\succ\in\bar{\mathcal{R}}^t}$, and $\bar{A}^{t*}$ be the matrix constructed from $\bar{A}^t$ by removing the rows that correspond to the last patch in every budget but the first one.

Lemma 4. *$\bar{A}^{t*}$ has full row rank.*

Proof. Take $\mathcal{T} = \{t\}$. By Corollary 2 in Saito (2017) or Theorem 2 in Dogan and Yildiz (2022), for any $\bar{\rho} \geq 0$ such that the sum over patches in any budget is equal to 1, there exists ν such that

$$\bar{A}^t \nu = \bar{\rho}.$$

Since $\bar{A}^t \alpha\nu = \alpha\bar{A}^t\nu = \alpha\bar{\rho}$ for any $\alpha \in \mathbb{R}$, $\bar{A}^t\nu$ can be any positive or any negative vector such that the sum over all patches in each budget does not depend on a budget (i.e. satisfies stability). Moreover, since any vector can be written as the sum of a positive and a negative vector, $\bar{A}^t\nu$ can represent any $\bar{\rho}$ such that sums over budgets are budget independent. Hence, if we remove the last row from every budget except the first one, we obtain that for any vector

$\bar{\rho}^*$, there exists ν such that $\bar{A}^{t*}\nu = \bar{\rho}^*$. Thus, $\bar{A}^{t*}$ is of full row rank. Indeed, if it was not, then there would exist $\xi \neq 0$ such that $\xi' \bar{A}^{t*}\nu = 0 \cdot \nu = 0$ for all ν. Therefore, we would have $\xi' \bar{A}^{t*}\nu = \xi' \bar{\rho}^* = 0$ for all $\bar{\rho}^*$. This can only be true if $\xi = 0$, contradicting $\xi \neq 0$. ∎

10.6. Proof of Theorem 4

Necessity. Suppose that ρ is rationalized by DRUM.

Necessity of stability. Follows from Theorem 3.

Necessity of D-monotonicity. Note that $A_T = A^1 \otimes A_{T-1}$, where

$$A^1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

Note that, since in every time period there are only 2 budgets, if ρ is rationalized by DRUM, then by Theorem 1, there exists component-wise nonnegative ν (i.e. $\nu \geq 0$) such that $A_T\nu = \rho$. We next show that this $\nu \geq 0$ together with stability, which we already showed to be satisfied, implies D-monotonicity.

First, note that we can partition ν into 3 vectors (ν_1^1, ν_2^1, and ν_3^1) and ρ into 4 vectors ($\rho_{1|1}^1$, $\rho_{2|1}^1$, $\rho_{1|2}^1$, and $\rho_{2|2}^1$) such that

$$\begin{pmatrix} \rho_{1|1}^1 \\ \rho_{2|1}^1 \\ \rho_{1|2}^1 \\ \rho_{2|2}^1 \end{pmatrix} = \rho = A_T\nu = A_1 \otimes A_{T-1}\nu = \begin{pmatrix} A_{T-1} & A_{T-1} & 0 \\ 0 & 0 & A_{T-1} \\ A_{T-1} & 0 & 0 \\ 0 & A_{T-1} & A_{T-1} \end{pmatrix} \begin{pmatrix} \nu_1^1 \\ \nu_2^1 \\ \nu_3^1 \end{pmatrix} = \begin{pmatrix} A_{T-1}(\nu_1^1 + \nu_2^1) \\ A_{T-1}\nu_3^1 \\ A_{T-1}\nu_1^1 \\ A_{T-1}(\nu_2^1 + \nu_3^1) \end{pmatrix}.$$

In this representation, $\rho_{i|j}^1$ correspond to all choice paths that contain patch $x_{i|j}^1$. Subtracting the third line from the first one, and the second line from the fourth one in the last system of equations, we obtain that

$$\rho^1_{1|1} - \rho^1_{1|2} = \rho^1_{2|2} - \rho^1_{2|1} = A_{T-1}\nu^1_2 \geq 0,$$

where the last inequality follows from $\nu \geq 0$ and A_{T-1} consisting of zeros and ones. Thus, $\mathrm{D}\left(x^1_{i'_1|j'_1}\right)\left[\rho(x_{\mathbf{i}|\mathbf{j}})\right] \geq 0$ if $x^1_{i'_1|j'_1} >^D x^1_{i_1|j_1}$.

Applying the above arguments to $\rho^1_{1|1} - \rho^1_{1|2} = A_{T-1}\nu^1_2$, we obtain that

$$(\rho^1_{1|1,1|1} - \rho^1_{1|2,1|1}) - (\rho^1_{1|1,1|2} - \rho^1_{1|2,1|2}) = (\rho^1_{2|2,2|2} - \rho^1_{2|1,2|2}) - (\rho^1_{2|2,2|1} - \rho^1_{2|1,2|1}) = A_{T-2}\nu^2_2 \geq 0,$$

where $\rho^1_{i|j,i'|j'}$ corresponds to all choice paths that contain patches $x^1_{i|j}$ and $x^2_{i'|j'}$. Thus, $\mathrm{D}\left(x^2_{i'_2|j'_2}\right)\mathrm{D}\left(x^1_{i'_1|j'_1}\right)\left[\rho(x_{\mathbf{i}|\mathbf{j}})\right] \geq 0$ if $x^1_{i'_1|j'_1} >^D x^1_{i_1|j_1}$ and $x^2_{i'_2|j'_2} >^D x^2_{i_2|j_2}$. Repeating these steps we can get that for all $K \leq T$

$$\mathrm{D}\left(x^K_{i'_K|j'_K}\right)\ldots\mathrm{D}\left(x^2_{i'_2|j'_2}\right)\mathrm{D}\left(x^1_{i'_1|j'_1}\right)\left[\rho(x_{\mathbf{i}|\mathbf{j}})\right] \geq 0$$

if $x^t_{i'_t|j'_t} >^D x^t_{i_t|j_t}$ for all $t = 1, \ldots, K$. Note that for any permutation of time periods the matrix A_T does not change. Hence, the above steps can be performed for any permutation of $x_{\mathbf{i}|\mathbf{j}}$ and D-monotonicity is satisfied.

Sufficiency. Assume that ρ is stable and D-monotone. Define $H_L = H^1 \otimes H_{L-1}$ and $P_{A_T} = P_{A_1} \otimes P_{A_{T-1}}$, where

$$H_1 = (A'_1 A_1)^{-1} A'_1 = \begin{pmatrix} 0.25 & 0.25 & 0.75 & -0.25 \\ 0.5 & -0.5 & -0.5 & 0.5 \\ -0.25 & 0.75 & 0.25 & 0.25 \end{pmatrix}$$

and

$$P_{A_1} = A_1(A_1'A_1)^{-1}A_1' = \begin{pmatrix} 0.75 & -0.25 & 0.25 & 0.25 \\ -0.25 & 0.75 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.75 & -0.25 \\ 0.25 & 0.25 & -0.25 & 0.75 \end{pmatrix}.$$

If we show that $\nu = H_T\rho$ satisfies (i) $\nu \geq 0$ and (ii) $A_T\nu = \rho$, then by Theorem 1 ρ is rationalized by DRUM.

Step 1: $\nu \geq 0$. Note that

$$\begin{pmatrix} \nu_1^1 \\ \nu_2^1 \\ \nu_3^1 \end{pmatrix} = \nu = H_T\rho = H_1 \otimes H_{T-1}\rho = \begin{pmatrix} 0.25H_{T-1} & 0.25H_{T-1} & 0.75H_{T-1} & -0.25H_{T-1} \\ 0.5H_{T-1} & -0.5H_{T-1} & -0.5H_{T-1} & 0.5H_{T-1} \\ -0.25H_{T-1} & 0.75H_{T-1} & 0.25H_{T-1} & 0.25H_{T-1} \end{pmatrix} \begin{pmatrix} \rho_{1|1}^1 \\ \rho_{2|1}^1 \\ \rho_{1|2}^1 \\ \rho_{2|2}^1 \end{pmatrix}.$$

Applying stability (i.e., $\rho_{1|2}^1 + \rho_{2|2}^1 = \rho_{1|1}^1 + \rho_{2|1}^1$), we can conclude that

$$\begin{pmatrix} \nu_1^1 \\ \nu_2^1 \\ \nu_3^1 \end{pmatrix} = \begin{pmatrix} H_{T-1}\rho_{1|2}^1 \\ H_{T-1}(\rho_{1|1}^1 - \rho_{1|2}^1) \\ H_{T-1}\rho_{2|1}^1 \end{pmatrix}.$$

If we next apply the above steps to $\nu_1^1 = H_{T-1}\rho_{1|2}^1$, then we can obtain that

$$\begin{pmatrix} \nu_{11}^1 \\ \nu_{12}^1 \\ \nu_{13}^1 \end{pmatrix} = \begin{pmatrix} H_{T-2}\rho_{1|2,1|2}^1 \\ H_{T-2}(\rho_{1|1,1|2}^1 - \rho_{1|2,1|2}^1) \\ H_{T-2}\rho_{1|2,2|1}^1 \end{pmatrix},$$

where $\rho_{i|j,i'|j'}^1$ corresponds to all choice paths that contain patches $x_{i|j}^1$ and $x_{i'|j'}^2$. If, instead, we apply it to $\nu_2^1 = H_{T-1}(\rho^1_{1|1} - \rho^1_{1|2})$, then we obtain

$$\begin{pmatrix} \nu^1_{21} \\ \nu^1_{22} \\ \nu^1_{23} \end{pmatrix} = \begin{pmatrix} H_{T-2}(\rho^1_{1|1,1|2} - \rho^1_{1|2,1|2}) \\ H_{T-2}((\rho^1_{1|1,1|1} - \rho^1_{1|2,1|1}) - (\rho^1_{1|1,1|2} - \rho^1_{1|2,1|2})) \\ H_{T-2}(\rho^1_{1|1,2|1} - \rho^1_{1|2,2|1}) \end{pmatrix}.$$

Repeating the above steps T times, we obtain that every component of ν is either equal to $\rho\left(\left(x^t_{1|2}\right)_{t\in\mathcal{T}}\right) \geq 0$, or $\rho\left(\left(x^t_{2|1}\right)_{t\in\mathcal{T}}\right) \geq 0$, or

$$\mathrm{D}\left(x^{\mathbf{t}}_{\mathbf{i}'|\mathbf{j}'}\right)\left[\rho\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] \geq 0,$$

for some $\mathbf{t} \in \boldsymbol{\mathcal{T}}$ and some $x_{\mathbf{i}|\mathbf{j}}$. The last inequality follows from $x^t_{i'_t|j'_t} >^D x^t_{i_t|j_t}$ for all $t \in \mathbf{t}$ and D-monotonicity. Hence, the proposed ν is nonnegative.

Step 2: $A_T\nu = \rho$. Note that

$$A_T\nu = P_{A_T}\rho = \begin{pmatrix} 0.75P_{A_{T-1}} & -0.25P_{A_{T-1}} & 0.25P_{A_{T-1}} & 0.25P_{A_{T-1}} \\ -0.25P_{A_{T-1}} & 0.75P_{A_{T-1}} & 0.25P_{A_{T-1}} & 0.25P_{A_{T-1}} \\ 0.25P_{A_{T-1}} & 0.25P_{A_{T-1}} & 0.75P_{A_{T-1}} & -0.25P_{A_{T-1}} \\ 0.25P_{A_{T-1}} & 0.25P_{A_{T-1}} & -0.25P_{A_{T-1}} & 0.75P_{A_{T-1}} \end{pmatrix} \begin{pmatrix} \rho^1_{1|1} \\ \rho^1_{2|1} \\ \rho^1_{1|2} \\ \rho^1_{2|2} \end{pmatrix}.$$

Since stability implies that $\rho^1_{1|1} + \rho^1_{2|1} = \rho^1_{1|2} + \rho^1_{2|2}$, we obtain

$$A_T\nu = P_{A_T}\rho = \begin{pmatrix} P_{A_{T-1}}\rho^1_{1|1} \\ P_{A_{T-1}}\rho^1_{2|1} \\ P_{A_{T-1}}\rho^1_{1|2} \\ P_{A_{T-1}}\rho^1_{2|2} \end{pmatrix}.$$

Repeating the above step one more time we obtain that

$$P_{A_{T-1}}\rho^1_{i|j} = \begin{pmatrix} P_{A_{T-2}}\rho^1_{i|j,1|1} \\ P_{A_{T-2}}\rho^1_{i|j,2|1} \\ P_{A_{T-2}}\rho^1_{i|j,1|2} \\ P_{A_{T-2}}\rho^1_{i|j,2|2} \end{pmatrix},$$

where $i, j \in \{1, 2\}$. Repeating the above steps T times for each subvector, we obtain

$$P_{A_T}\rho = \rho.$$

Hence, $A\nu = \rho$.

10.7. Proof of Theorem 5

Proof. (i) implies (ii). If ρ is consistent with RUM, then there exists an increasing random utility function u^t distributed according to μ such that $\mu(\arg\max_{y\in B^t_{j_t}} u^t(y) = x_{i_t|j_t}) = \rho(x_{i_t|j_t})$ for all $j_t \in \mathcal{J}^t$ and $i_t \in \mathcal{I}^t_{j_t}$. Using this random u^t we can extend ρ to $\bar{\mathbf{J}}^t$, so the BM inequalities are satisfied and the constructed $\bar{\rho}$ agrees with ρ. It is left to show that $\bar{\rho}$ is IU-consistent. Towards a contradiction, assume that there exists a menu, $B^t_{j_t}$, and $x_{i_t|j_t}$ in it such that $\bar{\rho}(x_{i_t|j_t}) > 0$, and that for all $x_{i_t|j_t}$ there exists some $S \subseteq B^t_{j_t}$ such that $S >^t x_{i_t|j_t}$. This is impossible since u^t is assumed to be a monotone function on $>^t$, so no monotone function would choose a point in $x_{i_t|j_t}$ when better points are available in other patches. This contradiction completes the proof.

(ii) implies (i). Let $\bar{\mathcal{R}}^t$ be the set of linear orders on X^t. By the result in Falmagne (1978), we know that there is a $\nu \in \Delta\left(\bar{\mathcal{R}}^t\right)$ such that

$$\overline{\rho}(x_{\mathbf{i}|\mathbf{j}}) = \sum_{\succ\in\bar{\mathcal{R}}^t} \nu(\succ)\mathbb{1}\left(x_{\mathbf{i}|\mathbf{j}} \succ y_{\mathbf{i}|\mathbf{j}} \quad \forall y_{\mathbf{i}|\mathbf{j}}, \mathbf{i} \in \mathbf{I}_{\mathbf{j}}\right).$$

It only remains to show that the stochastic demand generated by the mixture of linear orders on the extended set of menus assigns a zero measure to linear orders that are not extensions of $>^t$. Since ρ is IU-consistent, $\nu(\succ) = 0$ for any $\succ \in \bar{\mathcal{R}}^t$ that is not an extension of the order $>^t$. To show this is true, we prove the contrapositive. Namely, if $\nu(\succ) > 0$ for some $\succ \in \bar{\mathcal{R}}^t$ that is not an extension of the order $>^t$, then there exist $S, S' \subseteq \mathbf{X}^t$ such that $S >^t S'$ yet there is a $y \in S'$ such that $y \succ x$ for all $x \in S$. Thus, IU-consistency fails for the virtual budget $\{y, x\}$.

(ii) is equivalent to (iii). The statement follows from the definition of matrix $\bar{H}^t$. ∎

10.8. Proof of Theorem 7

Proof. (i) implies (ii). Direct from arguments analogous to those made in Theorem 5, Theorem 1, and Theorem 3.

(ii) implies (i). We break the proof into two steps.

First step. Let $\bar{\mathcal{R}}$ be the set of linear order profiles in $\times_{t\in\mathcal{T}}\mathbf{X}^t$, with typical element $(\succ^t)_{t\in\mathcal{T}}$. For any $\bar{\rho}$ such that satisfy $\bar{\rho} \in \bigcap_{k_1,\cdots,k_T\geq 1}\left\{\Gamma_{\mathbf{k}}^{\phi^*\prime} z \;:\; \left(\otimes_{t\in\mathcal{T}}\bar{H}^{t,\otimes_{k_t}}\right) z \geq 0\right\}$ and stability, we can use the results in Theorem 2 and the results in Theorem 3, to ensure that there exists a $\nu \in \Delta(\bar{\mathcal{R}})$ such that

$$\bar{\rho}(x_{\mathbf{i}|\mathbf{j}}) = \sum_{(\succ^t)_{t\in\mathcal{T}}\in\mathcal{R}^*} \nu((\succ^t)_{t\in\mathcal{T}})\mathbb{1}\left(x^t_{i_t|j_t} \succ^t y \quad \forall y \in B^t_{j_t} \quad \forall t \in \mathcal{T}\right).$$

Since ρ is IU-consistent, $\nu((\succ^t)_{t\in\mathcal{T}}) = 0$ for any $(\succ^t)_{t\in\mathcal{T}} \in \bar{\mathcal{R}}$ that contains some element $\succ^t$ that is not an extension of the order $>^t$. To show this is true, we prove the contrapositive. Namely, if $\nu((\succ^t)_{t\in\mathcal{T}}) > 0$ for some $(\succ^t)_{t\in\mathcal{T}} \in \bar{\mathcal{R}}$ that is not an extension of the order $>^t$, then there exist nonempty $S, S' \subseteq \mathbf{X}^t$ such that $S >^t S'$ yet for an element $y \in S'$ $y \succ x$ for all $x \in S$. Thus, IU-consistency fails for the virtual budget path that contains the budget $\{y, x\}$, for any selection of $x \in S$, at time t. ∎

10.9. Proof of Proposition 2

Proof. We provide here the proof of Proposition 2. Assume towards a contradiction that ρ is rationalized by DRUM with μ that satisfies constancy and SARPD is violated for some $\mathbf{j}$. Hence, there exist some y^{t_1}, y^{t_N}, and some $u \in U$ such that $u\left(y^{t_1}_{i_t|j_t}\right) > u\left(x^{t_N}_{i_{t_n}|j_{t_n}}\right)$. However, the violation of SARPD implies that $u\left(x^{t_1}_{i_{t_1}|j_{t_1}}\right) > u\left(x^{t_1}_{i_{t_1}|j_{t_1}}\right)$ which is impossible. SARPD rules out the possibility that some individuals in the population violate SARP. When constancy is relaxed, we need to obtain cross-sectional variation (i.e., more than one budget path) to test DRUM. ■

10.10. Proof of Proposition 3

Proof. Let, $\rho^{-t}((x^{\tau}_{i_\tau|j_\tau})_{\tau\in\mathcal{T}\setminus\{t\}}) = \left(\sum_{i\in\mathcal{I}^t_j} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right)\right)$. We also define the vector

$$\rho^{-\tau} = (\rho((x^{t}_{i_t|j_t})_{t\in\mathcal{T}\setminus\{\tau\}}))_{\mathbf{j}\in\mathbf{J},\mathbf{i}\in\mathbf{I}_j}.$$

Note that ρ^{-1} is of the same length that $\rho^1_{i|j}$ for any patch $x^1_{i|j}$. We let $\mathcal{R}_t$ be the set of linear orders at time $t \in \mathcal{T}$. The scalar a_{t,r_t,i_k,j_k} is the entry of matrix A_t for column corresponding to r_t and row corresponding to i_k, j_k.

Lemma 5. *If the vector representation of P, ρ, is consistent with DRUM, then for every finite sequence of patches (including repetitions), k, $\{(i_k, j_k)\}$ such that $j_k \in \mathcal{J}^t$ and $i_k \in \mathcal{I}^t_{j_k}$*

$$\sum_k \rho^1_{i_k|j_k} \leq \rho^{-1} \max_{r_t\in\mathcal{R}_t} \sum_k a_{t,r_t,i_k,j_k}.$$

The condition above implies the fact that marginals, conditionals are consistent with RUM. Assume that ρ is interior (i.e., rule out zero probabilities on choice paths), then the condition above implies that the marginal probability

$$\rho((x^t_{i_t|j_t})_{t\in\mathcal{T}})|(x^\tau_{i_\tau|j_\tau})_{\tau\in\mathcal{T}\setminus\{1\}}) = \frac{\rho((x^t_{i_t|j_t})_{t\in\mathcal{T}})}{\rho((x^\tau_{i_\tau|j_\tau})_{\tau\in\mathcal{T}\setminus\{1\}})},$$

is consistent with (static) RUM. In that case the condition above is just the ASRP of McFadden and Richter (1990). It is easy to see that the same reasoning can be done recursively and for any permutation of time, so all conditional probabilities of choice, as defined above, are consistent with (static) RUM if the vector representation ρ is consistent with DRUM. ■