2023

2023-2 Dynamic and Stochastic Rational Behavior

Nail Kashaev

Charles Gauthier

Victor H. Aguiar

Dynamic and Stochasitc Rational Behavior

by

Nail Kashaev, Charles Gauthier, and Victor H. Aguiar

Research Report # 2023-2 February 2023

Department of Economics

Research Report Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

Dynamic and Stochastic Rational Behavior*

Nail Kashaev ⓡ Charles Gauthier ⓡ Victor H. Aguiar†

February 8, 2023

Abstract We analyze consumer demand behavior using Dynamic Random Utility Model (DRUM). Under DRUM, a consumer draws a utility function from a stochastic utility process in each period and maximizes this utility subject to her budget constraint. DRUM allows unrestricted time correlation and cross-section heterogeneity in preferences. We fully characterize DRUM for a panel data of consumer choices and budgets. DRUM is linked to a finite mixture of deterministic behavior represented as the Kronecker product of static rationalizable behavior. We provide 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 for time series and more informative than RUM. We show the feasibility of the statistical test of DRUM in a Monte Carlo study.

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, 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.

†Kashaev: Department of Economics, University of Western Ontario; nkashaev@uwo.ca. Gauthier: ECARES, Université Libre de Bruxelles; charles.gauthier@ulb.be. Aguiar: Department of Economics, University of Western Ontario; vaguiar@uwo.ca.

1. Introduction

One key question in economics is whether consumer behavior is rational. 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 consumer behavior that is stochastic and dynamic–Dynamic Random Utility Model (DRUM). Under DRUM, each consumer at each time maximizes the realized utility from a stochastic utility process subject to a budget constraint. We provide a revealed preference characterization of DRUM when the longitudinal distribution of demand is observed for a finite collection of 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: Afriat (1967)'s framework of static utility maximization and McFadden-Richter (1990)'s framework, called random utility model (RUM). DRUM addresses several important empirical limitations of these models. In particular, Afriat's framework is under scrutiny due to experimental and field evidence against it.[1] There is evidence that failures of 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). DRUM allows preferences to change freely in time. In contrast to Afriat's framework, RUM has found reasonable success explaining repeated cross-sections of household choices (Kawaguchi, 2017, Kitamura and

[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 violations of rationality originally were 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).

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

Our first revealed preference characterization of DRUM is analogous to the RUM characterization in McFadden-Richter's work. We exploit the fact that DRUM is associated with a finite mixture of demand profiles in time. We obtain results analogous to Kitamura and Stoye (2018) (henceforth KS) and McFadden and Richter (1990), Kawaguchi (2017) with a dynamic version of the Axiom of Stochastic Revealed Preferences. This characterization lends itself to statistical testing using results in KS. Also, our 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 find the mixture representation of DRUM can be obtained using a Kronecker product of the mixture representation of RUM for each period.[2] This observation is vital to obtain: (i) computational gains for testing because of the modularity of the mixture representation that is parallelizable; and (ii) a recursive characterization of DRUM. To provide this characterization of DRUM we prove a generalization of the Weyl-Minkowski theorem for cones that exploits the recursive structure of the DRUM induced by the Kronecker product. 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 was the first to observe that the empirical content of RUM can be expressed as cone restrictions. In particular, they notice that the $\mathcal{H}$-representation corresponds to what decision theorists would call an axiomatic characterization of RUM.[3] This new mathematical result enables the conversion of static

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

[3]In fact, we notice that Block Marschak inequalities are the $\mathcal{H}$-representation of RUM in abstract setups with finite discrete choice.

RUM characterizations to their dynamic analogous using the Kronecker product structure. We use this result to provide a novel characterization of DRUM using a recursive version of Block and Marschak (1960) inequalities. We believe that our generalized Weyl-Minkowski theorem can be helpful beyond DRUM (e.g., models of bounded rationality such as random consideration (Cattaneo, Ma, Masatlioglu and Suleymanov, 2020) can be extended in the spirit of DRUM).

The generalized Weyl-Minkowski theorem enables us to provide a novel behavioral condition that is necessary for consistency of the longitudinal distribution of demand with DRUM (D-monotonicity). It is also sufficient in a simple-setup: (i) for any finite number of goods and 2 budgets per time period, and (ii) 2 goods and finitely many budgets per time 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, Afriat's and McFadden-Richter's frameworks. Afriat's framework requires observing a *time-series* of choices and budgets of a given consumer and assumes that a consumer maximizes the same utility function each time. When this assumption about the utility stochastic process being constant over time is relaxed, there are no empirical implications when observing only a time-series of choices. However, using a *panel*, DRUM bounds the share of consumers whose choices contain a revealed preference violation in the Afriat's sense. RUM instead requires observing a *cross-section* of choices and budgets of a population of consumers. There is no time dimension in RUM. One can ignore the panel structure, but unfortunately, this approach misses the potential temporal correlation of utilities. As a result, there are certain panels of choices over budgets that, when marginalized, are consistent with RUM, but not with

DRUM. In other words, ignoring the time dimension of choice may lead to false positives when testing RUM. Importantly, our setup keeps the same key assumption in McFadden-Richter's framework. Namely, that the distribution over utilities in time does not depend on the budgets that the consumer faces in time.[4]

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 the constant preferences across time assumption in 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).[5]

The DRUM framework is rich and extends well beyond the Afriat's 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)) for choices over portfolios of securities as in Polisson, Quah and Renou (2020); (iii) static utility maximization in a population (without measurement error) (Aguiar and Kashaev, 2021); (iv) dynamic utility maximization in a population[6] (Browning, 1989, Gauthier, 2018, Aguiar and Kashaev, 2021); (v) changing utility or multiple-selves models (Cherchye *et al.*, 2017); and changing-taste modeled with a constant utility in time with an additive shock (Adams, Blundell, Browning and Crawford, 2015).

DRUM was first defined in Strzalecki (2021) in an abstract domain for discrete choice. Frick *et al.* (2019) provide an axiomatic characterization of it for a rich domain with decision trees and an expected utility restriction on the stochastic utility process. We provide the first

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

[5]In practice, panels of choices are often pooled in the time dimension to create a cross-section with sufficient variation of budgets (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.

[6]This requires a redefinition of price to be an effective price that includes an adjustment due to interests rates as described in Aguiar and Kashaev (2021).

characterization of DRUM for a consumer choice domain with limited observability on budgets without requiring any restriction on preferences.

Related Literature. Recent interest in DRUM in finite abstract discrete choice space has provided partial characterizations of it when the primitive is the joint distribution of choices across time and with total menu variation. Li (2021) provides an axiomatic characterization of DRUM (analogous to Block-Marschak (BM) inequalities for RUM Block and Marschak, 1960) for any finite number of time periods, full menu variation but for the cardinality of the choice set less than or equal to 3. Chambers, Masatlioglu and Turansick (2021) consider correlated choice, which is the joint distribution of choice on a pair of menus; the choice may be made by a group instead of a single decision-maker. Some versions of this model can be considered dynamic choices when a decision-maker's multiple selves are making decisions. They did not characterize the problem for the general case with arbitrary cardinality of the choice set and an arbitrary number of selves. Note that the primitive in both Li (2021) and Chambers *et al.* (2021) differs from ours in the general setup. Importantly, in our setup, the domain of classical consumer choice is endowed with a primitive order (i.e., the vector order), and preference revelation respects that primitive order. Our DRUM will respect this primitive order and restrict utilities to be monotone. Another difference is that we deal with a continuum of choices and limited observability of menus and histories. Finally, Li (2021) and Chambers *et al.* (2021) assume comprehensive menu variation that allows them to provide a BM-like characterization for special cases of their setup, exploiting the nested structure of menus under the set containment. In contrast, choice sets in our setup are not nested, so we cannot use the characterizations in Li (2021) and Chambers *et al.* (2021).

We fully characterize DRUM in the abstract setup of Li (2021) and Chambers *et al.* (2021) as a byproduct of our investigation for the general case of finite abstract discrete choice with full menu variation. We provide a form of recursive BM inequalities that works for cases of limited observability of menus/choice sets. This characterization is an application of our new generalization of the Weyl-Minkowski theorem of convex cones to our setup.

We contribute to the literature that studies random exponential discounting as in Browning (1989) for the demand setup generalizing the setup of Deb *et al.* (2021) to a dynamic setup. Apesteguia, Ballester and Gutierrez-Daza (2022) introduces a heterogeneous risk and time preferences model with exponential discounting and time separability. The choice domains studied there differ from our setup. Also, their setup is semiparametric, whereas ours is nonparametric. Lu and Saito (2018) also studies exponential discounting with a random discount factor. Choices over consumption streams are made in the first period stochastically.

Aguiar and Kashaev (2021) studies a panel setup but uses a first-order-conditions approach to deal with some forms of dynamic preferences. Mainly, they allow measurement error, which can be mapped to trembling-hand or misperception errors. However, their setup does not allow for changing utility beyond a changing discount factor or marginal utility of income. Im and Rehbeck (2021) study the McFadden-Richter's framework and its inability to use a panel structure. However, they propose to check individual static rationality, like in Afriat's framework, as a potential solution. Here, we generalize Afriat's framework to allow an individual's utility to change over time while exploiting the panel structure to obtain more empirical implications than McFadden-Richter's framework.

Outine. The paper is organized as follows, Section 2 introduces the setup. Section 3 provides a characterization of DRUM. In Section 4, we study a simple setup with two budgets in each time period. Section 5 presents a BM-like representation for DRUM, to do so it provides a new recursive characterization of convex cones with a tensor structure. Section 6 provides an extension of the main model that allows endogenous expenditure. Section 7 provides a synthesis of Afriat's and McFadden-Richter's setups. Section 8 provides results about dynamic counterfactual analysis. Section 10 concludes. All proofs can be found in Appendix 11.

2. Setup

Let $X \subseteq \mathbb{R}^K_+$ be the consumption space with finite $K \geq 2$ goods.[7] We consider a time window $\mathcal{T} = \{1, \cdots, T\}$ with a finite terminal period $T \geq 1$. 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.

Define a budget path as an ordered collection of indexes $\mathbf{j} = (j_t)_{t \in \mathcal{T}}$, $j_t \in \mathcal{J}^t$. Budget paths encode budgets that were faced by agents in different time periods. 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 the Cartesian product of T repetitions of X, $\times_{t \in \mathcal{T}} X$. The primitive in our 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*.

Some examples of datasets where a dynamic stochastic demand system is (partially) observed are: (i) household longitudinal survey datasets (e.g., *Encuesta de Presupuestos Familiares* in Spain and Progresa Household Survey in Mexico, Deb *et al.*, 2021, Aguiar and Kashaev, 2021); (ii) scanner datasets (e.g., Nielsen homescan data, Gauthier, 2018); and (iii) experimental datasets where subjects need to pick a point on the budget line several times (e.g., experiments on preferences over giving as in Porter and Adams, 2016). In survey datasets, information about household purchases is usually collected several times a year (e.g., quarterly). For a given time period, budget variation across households is driven by spatial (e.g. regional) price variation (Aguiar and Kashaev, 2021). Scanner datasets 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, often, each subject faces at

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

random a budget path drawn from the same set of budgets for all subjects. Since the number of subjects is usually much bigger than the number of budget paths, there are many subjects facing the same budget path.

Given P, we can define a Dynamic Random Utility Model (DRUM). Let U denote the set of all continuous, strictly concave, and monotone utility functions that map X to $\mathbb{R}$ and $\mathcal{U} = \times_{t\in\mathcal{T}} U$ be the Cartesian product of T repetitions of U.

Definition 1 (DRUM)**.** The dynamic stochastic demand P is consistent with DRUM if there 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 $O^t \subseteq X$, $t \in \mathcal{T}$, where $u = (u^t)_{t\in\mathcal{T}}$.

When $T = 1$, DRUM coincides with RUM, where every agent maximizes her utility function u^1 over a budget set 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 the preference heterogeneity in cross-sections (i.e. across agents) and requires μ not to depend on the budget paths and the alternatives in the consumption space. 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). Thus, it does not restrict preferences of individuals in cross-sections. However, 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 7.

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 is such that $\alpha_t = \max\{\min\{\alpha_{t-1} + \epsilon_t, 1\}, 0\}$, where

ϵ_ts are independent identically distributed mean-zero random innovation with variance σ^2 and $\alpha_1 \in [0,1]$. Note that α_ts are correlated across time. If $\sigma^2 = 0$, α_t does not change over time, thus, the consumer share of her wealth in each good remains stable in time. 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 unction 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 exogeneity assumption is relaxed in Section 6. There we extend the main model to endogenize expenditure. This allows us to connect our setup with the possibility of savings and intertemporal consumption and discounting in the sense of Browning (1989). We provide this example below.

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, the discount factor δ, history of incomes captured by information 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

$$p_\tau^{'} c_\tau(y) + s_t(y) = y_\tau + (1 + r_\tau)s_{\tau-1}(y),$$

where the expectation is taken with respect to 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

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

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

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

Correlation in income across time would generate correlation between $\{\hat{v}^t\}_{t\in\mathcal{T}}$. We show in Section 6 that if we normalize consumption by the expenditure in each time period this example can be mapped to DRUM.[8] In addition, one has to make make the standard assumptions that different individuals have different u, δ, and y, and crucially that their joint distribution does not depend on prices. Normalization of consumption is important because without it this example is not consistent with DRUM because of the dependence of v^t on expenditure/income through savings.

3. Characterization of DRUM

Here we provide a characterization of rationalizability by DRUM when P is observed (estimable). The main result in this section will be an analogue of the McFadden-Richter's and

[8]This normalization is analogous to the one used in the static problem in Deb *et al.* (2021).

KS's results for RUM.

3.1. Patches

Monotonicity of the utility functions generates choices on the budget line. In the RUM 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 lines called *patches*.

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

Definition 2 (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 planes. A typical patch belongs to one budget plane. However, intersection patches always belong to several budget planes. 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.

Given a budget path $\mathbf{j} \in \mathbf{J}$, a *choice path* is an array of patches $x_{\mathbf{i}|\mathbf{j}} = \left(x^t_{i_t|j_t}\right)_{t \in \mathcal{T}}$ 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 budget path, a choice path encodes choices of an agent in a given sequence of budget sets she faced. The set of all possible choice path index sets $\mathbf{i}$, given the budget path $\mathbf{j}$, is denoted by $\mathbf{I}_\mathbf{j}$. Let

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

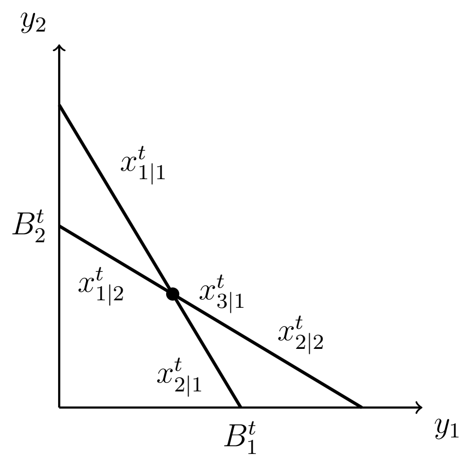

Figure 1 – Patches for the case with $K = 2$ goods and $J^t = 2$ budgets. The only intersection patch is $x^t_{3|1}$, which is the intersection of B^t_1 and B^t_2.

denote the fraction of agents who pick from a choice path $x_{\mathbf{i}|\mathbf{j}}$ given a budget path $\mathbf{j}$. Hence, $\rho\left(x_{\mathbf{i}|\mathbf{j}}\right) \geq 0$ and $\sum_{\mathbf{i}\in\mathbf{I_j}} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right) = 1$.

The main building block of our framework is *the vector representation* of P

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

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

Given the finite set of patches, let a preference profile be $\mathbf{r} = (r_t)_{\in\mathcal{T}}$, where r_t is a linear order defined on the set of patches available at time t, $\bigcup_{j\in\mathcal{J}^t, i\in\mathcal{I}^t_j} x^t_{i|j}$. Given the preference profile $\mathbf{r}$, we can encode choices in different time periods and budgets 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 patch $x^t_{i_t|j_t}$ is the best patch 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 profiles of preferences

$\mathbf{r}$ for which there exists $u_r = (u_r^t)_{t\in\mathcal{T}} \in \mathcal{U}$ such that

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

We form matrix A_T by stacking the column vectors $a_{\mathbf{r}}$ for all preference profiles in $\mathbf{r} \in \mathcal{R}$. 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 3 (Axiom of Dynamic Stochastic Revealed Preference, ADSRP)**.** A vector representation ρ satisfies ADSRP if for every finite sequence of pairs of budget 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) The dynamic stochastic demand P *is consistent with DRUM.*

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

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

(iv) The vector representation ρ *satisfies the ADSRP.*

The main part of the proof of Theorem 1 is based on the fact that, without loss of generality, P can be reduced to a demand that assigns mass only to the representative elements of patches (e.g., geometric centers) along a choice path. Then the equivalence of (i)-(iv) is analogous to proof 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, even if ones uses the tools of Smeulders, Cherchye and De Rock (2021), testing DRUM may seem impossible for even relatively small T. 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 (i.e., testing DRUM is not much harder than testing RUM).

Lemma 1. *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, where $A^1 \otimes A^2$ indicates the Kronecker product of matrices A^1 and A^2.*

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 budget 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 budgets where each budget 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 1 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 (or its approximation). In instances where the budget structure is such that $A^t = A^{t'}$ for $t \neq t'$ significant computational savings are achieved. Note that $A^t = A^{t'}$ can occur without budgets in t and t' being the same. In fact, the matrix A^t depends only on the intersection structure induced by the budgets at t and not on the specific prices. This will become more transparent as we provide examples in the next section. Lemma 1 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. We can parallelize the computation of A_T. This recursivity or modularity is heavily exploited to obtain new theoretical results in the next sections including a recursive characterization of DRUM.

Unfortunately, the DRUM characterization in Theorem 1 does not provide an intuitive understanding of the behavioral implications of DRUM. The same is also true of the current characterization of RUM in the demand setup. Nevertheless, the recursive or modular structure of DRUM is useful to take our characterization to data. In the next sections, we provide an intuitive characterization of DRUM for a setting with two budgets per time period and a general recursive formulation for settings with many time periods and many budgets. These characterizations demonstrate that DRUM provides additional implications in longitudinal data than those in RUM. Also, it will become apparent 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 P.

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

In this section, we illustrate our setup and Theorem 1 in the 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 $x_{1|1}^t, x_{2|1}^t, x_{1|2}^t$, and $x_{2|2}^t$ (see Figure 2 for a graphical representation of the case with $K = 2$ goods).[9] We call choice path configurations implied by these 4 patches the *simple-setup* choice paths. An example of a budget path for

[9] Formally, $x_{1|1}^t = \{y \in B_1^t : p_{2,t}'y > w_{2,t}\}$, $x_{2|1}^t = \{y \in B_1^t : p_{2,t}'y < w_{2,t}\}$, $x_{1|2}^t = \{y \in B_2^t : p_{1,t}'y < w_{1,t}\}$, and $x_{2|2}^t = \{y \in B_2^t : p_{1,t}'y > w_{1,t}\}$.

$T = 2$ is $(2, 1)$ (i.e. B_2^1 and B_1^2), an example of a choice path in this budget path is $\left(x^1_{1|2}, x^2_{1|1}\right)$. Conditional on the budget path, the total probability of all possible choice paths is equal to 1 (i.e., $\rho\left(\left(x^1_{1|2}, x^2_{1|1}\right)\right) + \rho\left(\left(x^1_{2|2}, x^2_{1|1}\right)\right) + \rho\left(\left(x^1_{1|2}, x^2_{2|1}\right)\right) + \rho\left(\left(x^1_{2|2}, x^2_{2|1}\right)\right) = 1$).

In this setup, there are 3 rational demand types per time period that are described in Table 1.[10] Each demand type $\theta^t_{i,j}$ picks i-th patch in budget B_1^t and j-th patch in budget B_2^t at time t.

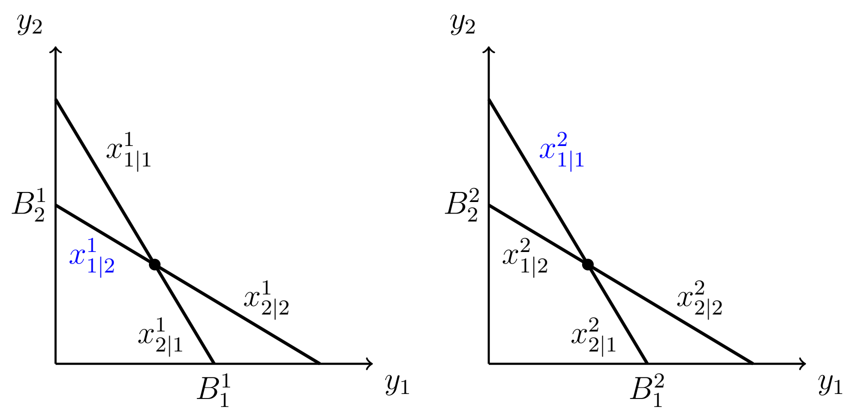

Figure 2 – Simple-setup for $K = 2$ goods and no intersection patches.

Type/Budget	B_1^t	B_2^t
$\theta^t_{1,1}$	$x^t_{1\|1}$	$x^t_{1\|2}$
$\theta^t_{1,2}$	$x^t_{1\|1}$	$x^t_{2\|2}$
$\theta^t_{2,2}$	$x^t_{2\|1}$	$x^t_{2\|2}$

Table 1 – Choices of 3 rational types in budgets B_1^t and B_2^t at time t.

Now we can write down the associated A_T matrix. Note that since, there are two intersecting budgets in every time period, $A^t = A^{t'}$ for all $t, t' \in \mathcal{T}$. Thus, by Lemma 1, we just need to compute the matrix A^t for one time period. The rows of this matrix correspond to the choice paths (4 possibles paths). We display the matrix A^t in Table 2 (for readability, we replace 0 by the symbol "$-$").

Using, A^t we can write down the matrix A_T for a time window of any size. For illustration purposes, we write it down below for $T = 2$ (i.e., $A_T = A^1 \otimes A^2$). In this case, since the

[10] The idea of writing demand types on patches was developed in KS and we use the convenient notation developed in Im and Rehbeck (2021).

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

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

	$(\theta^1_{1,1}, \theta^2_{1,1})$	$(\theta^1_{1,1}, \theta^2_{1,2})$	$(\theta^1_{1,1}, \theta^2_{2,2})$	$(\theta^1_{1,2}, \theta^2_{1,1})$	$(\theta^1_{1,2}, \theta^2_{1,2})$	$(\theta^1_{1,2}, \theta^2_{2,2})$	$(\theta^1_{2,2}, \theta^2_{1,1})$	$(\theta^1_{2,2}, \theta^2_{1,2})$	$(\theta^1_{2,2}, \theta^2_{2,2})$
$\left(x^1_{1\|1}, x^2_{1\|1}\right)$	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 3 – The matrix A_T for 2 time periods with 2 budgets per period. "$-$" corresponds to 0.

demand types correspond to a preference type, a demand profile $(\theta^1_{i,j}, \theta^2_{k,l})$ (i.e., $\theta^1_{i,j}$ in the first time period and $\theta^2_{k,l}$ in the second one) corresponds to a preference profile over the choice path (9 preference profiles). The rows of this matrix correspond to the choice paths (16 possibles paths).

4.1. D-monotonicity

In this section, we introduce a new behavioral restriction on ρ that is one of the two conditions that characterize DRUM in the simple-setup. We first introduce a static notion of dominance among patches.

Definition 4 (Patch-Revealed Dominance)**.** We say that patch $x^t_{i|j}$ is revealed dominant to $x^t_{i'|j'}$ or $x^t_{i_t|j_t} \succ^D x^t_{i'_t|j'_t}$ if for all $y \in x^t_{i|j}$ and $y' \in x^t_{i'|j'}$ (i) $p_{j,t}y > p_{j,t}y'$ and (ii) $p_{j',t}y > p_{j',t}y'$.

Patch-revealed dominance captures the idea 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 (in the Afriat's sense) to the elements of the dominant patch. We can visualize this ordering in Figure 2, where $x^1_{1|1} \succ^D x^1_{1|2}$ and $x^2_{2|2} \succ^D x^2_{2|1}$.

Let $x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^t_{i'_t|j'_t}$ denote a choice path where the t-th component of $x_{\mathbf{i}|\mathbf{j}}$ was replaced by $x^t_{i'_t|j'_t}$. We show that if ρ is consistent with DRUM and $x^t_{i'_t|j'_t} \succ^D x^t_{i_t|j_t}$ then

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

This monotonicity is intuitive, since DRUM implies that more DMs achieve their maximum at the *dominant* choice path.[11]

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

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

Thus, if there is a nonnegative solution of the system $A_T\nu = \rho$, then the following two inequalities have to be satisfied

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

These inequalities mean that the patches $x^1_{1|1}$ and $x^1_{2|2}$ should have a bigger probability mass than patches $x^1_{1|2}$ and $x^1_{2|1}$. Moreover, patch $x^1_{1|1}$ ($x^1_{2|2}$) dominates patch $x^1_{1|2}$ ($x^1_{2|1}$).

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

[11] It should be clear we can use the static patch dominance notion to order choice paths when they differ only in one patch in a fixed time period.

structure of the matrix A_T (recall that $A^1 = A^2$ in the simple-setup):

$$\rho = A_T\nu = A^1 \otimes A^2\nu = \begin{pmatrix} A^1 & A^1 & 0 \\ 0 & 0 & A^1 \\ A^1 & 0 & 0 \\ 0 & A^1 & A^1 \end{pmatrix} \begin{pmatrix} \nu_1^1 \\ \nu_2^1 \\ \nu_3^1 \end{pmatrix} = \begin{pmatrix} A^1(\nu_1^1 + \nu_2^1) \\ A^1\nu_3^1 \\ A^1\nu_1^1 \\ A^1(\nu_2^1 + \nu_3^1) \end{pmatrix}.$$

Similarly to the case with $T = 1$, since entries of A_1 are either 0 or 1, we can derive the following system of equations

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

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

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

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

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

Note that similar to the argument for the case of $T = 1$, it can be shown that

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

Thus, Inequality (1) 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 used in cases when $T > 2$. However, we would need to work with higher-order differences.

Next, we introduce the difference operator.

Definition 5 (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 $t_1 \neq t_2$

$$\begin{aligned}
&\mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\mathrm{D}\left(x^{t_1}_{i'_{t_1}|j'_{t_1}}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] = \mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\left[f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^{t_1}_{i'_{t_1}|j'_{t_1}}\right) - f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right] = \\
&\mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right) f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^{t_1}_{i'_{t_1}|j'_{t_1}}\right) - \mathrm{D}\left(x^{t_2}_{i'_{t_2}|j'_{t_2}}\right) f\left(x_{\mathbf{i}|\mathbf{j}}\right) t = \\
&\left[f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^{t_1}_{i'_{t_1}|j'_{t_1}} \downarrow_t x^{t_2}_{i'_{t_2}|j'_{t_2}}\right) - f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^{t_2}_{i'_{t_2}|j'_{t_2}}\right)\right] - \left[f\left(x_{\mathbf{i}|\mathbf{j}} \downarrow_t x^{t_1}_{i'_{t_1}|j'_{t_1}}\right) - f\left(x_{\mathbf{i}|\mathbf{j}}\right)\right],
\end{aligned}$$

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 the length of 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 6 (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 $x^t_{i'_t|i'_t} \succ^D x^t_{i_t|i_t}$ for all $t\in\mathbf{t}$

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

The notion of D-monotonicity builds a bridge between different time periods. It is a shape restriction on ρ since it captures the idea that as we switch to a dominant patch at one time period, the fraction of individuals being in this new choice path should increase. D-monotonicity is the generalization of the Weak Axiom of Stochastic Revealed Preference (WASRP) from the static case of RUM to our setup (Bandyopadhyay *et al.*, 1999). More precisely, since we impose monotonicity on the support of random utility process, for the static case, our condition coincides with the *stochastic substitutability* condition in Bandyopadhyay, Dasgupta and Pattanaik (2004).[12] For the case of two budgets, WASRP also a sufficient condition (Hoderlein and Stoye, 2014).[13] For the case of $T\geq 2$, D-monotonicity implies that the marginal impact in ρ of inserting a dominant patch in a given choice path is marginally increasing each time the D operator is applied to ρ. We highlight that Dasgupta and Pattanaik (2007) has shown that in the static case WASRP implies regularity, but it is not implied by it, when the domain of choices is not complete. This observation translates to the dynamic case as well.[14]

[12] Notice that in our setup we put zero measure on intersection patches whereas Bandyopadhyay *et al.* (2004) do not.

[13] D-monotonicity in the static case for the case of 2 goods, was shown to be also sufficient in Hoderlein and Stoye (2015).

[14] 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.

4.2. Stability

In this section, we introduce the second condition needed for the characterization of DRUM in the simple-setup.

Definition 7 (Stability)**.** We say that ρ is stable if $\sum_{i\in\mathcal{I}^t_j} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right)$ 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 in any t does not depend on the budget set in any other $t' \neq t$. Under stability, the marginal distribution of choices of consumers will not change due to the budget the consumers faced in the past or the budget 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. Stability was first defined in Strzalecki (2021).

4.3. Simple-setup characterization

We are ready to state our main result in this section.

Theorem 2. *For the simple-setup, the following are equivalent:*

(i) P *is rationalized by DRUM.*

(ii) ρ *is* D*-monotone and stable.*

Necessity is easy to verify. Sufficiency is proved constructively. Theorem 2 is not just a restatement of the Weyl-Minkowski Theorem, as stability and D-monotonicity correspond to the explicit H-representation of the cone restrictions implied by $A\nu = \rho$ for some $\nu \in \mathbb{R}^{|\mathcal{R}|_+}$ (V-representation). The H-representation is obtained by direct computation and can be directly used for testing DRUM in the simple setup. This characterization also provides the reader with a helpful intuition about the empirical content of DRUM.

	$x^2_{1\|1}$	$x^2_{2\|1}$	$x^2_{1\|2}$	$x^2_{2\|2}$
$x^1_{1\|1}$	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 4 – Matrix representation of ρ for $T = 2$ that violates D-monotonicity, but satisfies simple stability. "-" corresponds to 0

Next, we provide an example of ρ that violates D-monotonicity, but it satisfies stability. This example shows the two conditions are logically independent.

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

Another example of a stochastic demand that fails both conditions of the simple setup is discussed in Section 7.

We conclude this section by stating a by-product of the proof of 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 $\nu = \nu'$.*

Corollary 1 means DRUM has a uniqueness property at the ordinal level for the simple-setup. This result may not hold in settings with 3 or more budgets, even for one time period (see, Example 3.2 in KS).

5. $\mathcal{H}$-Representation of the General Model and Recursive Characterization of DRUM

In this section, we go back to the general primitive and study the necessity of suitable generalizations of the restrictions on behavior introduced in the simple-setup for consistency with DRUM. Moreover, we show how to convert the axiomatic characterization of (static) RUM (when they correspond to cone constraints) to a recursive characterization of DRUM. We end up the section by providing a characterization of RUM via recursive BM inequalities.

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

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

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

This is the so called $\mathcal{V}$-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:

$$\{t \,:\, B_T t \geq 0\}\,.$$

The $\mathcal{V}$-representation states that the observed distribution over patches is a finite mixture of deterministic types. Construction of matrix A_T is a straight-forward but computationally demanding problem (Kitamura and Stoye, 2018, Smeulders *et al.*, 2021).[15] Unfortunately, the $\mathcal{V}$-representation does not give any direct restrictions on the observed ρ. As a result, one can not derive any helpful intuition about the empirical content of DRUM using the $\mathcal{V}$-representation. In contrast, the $\mathcal{H}$-representation delivers direct, easy-to-test, tractable, and intuitive restrictions on the data (see Section 4). But construction of B_T from A_T is a nontrivial task that may become computationally intractable even for moderate T since the number of columns of A_T grows exponentially with T. In Lemma 1, we showed that one

[15] 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. Our insight is done for the new DRUM but its inspired by their observation.

could use the recursive structure of A_T (i.e., $A_T = A_{T-1} \otimes A^T$) to simplify its construction substantially. In this section, we show that the same intuition carries over to the construction of B_T: one can move from the $\mathcal{H}$-representation of RUM to the $\mathcal{H}$-representation of DRUM with almost no computational cost.

Our next result generalizes the Weyl-Minkowski theorem in a direction that is useful for our recursive setup.

Theorem 3. *If*

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

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

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

If, moreover, K^t has full row rank for all t, then these sets are equal.

To the best of our knowledge, Theorem 3 is a new result in convex analysis and of possible independent interest beyond economics. It allows one to easily construct $\mathcal{H}$-representations of cones generated by matrices constructed using Kronecker products from $\mathcal{H}$-representations of their one-time counterparts. Unfortunately, this theorem cannot be directly applied to our setting.

5.2. $\mathcal{H}$-representation of A_T

Theorem 3 gives necessary conditions for building the $\mathcal{H}$-representation of DRUM from its static components (i.e. $K^t = A^t$ and $L^t = B^t$). However, since A^t is never of full row rank, we can not establish full equivalence using Theorem 3. The rank is not full because of the "adding-up" constraint–one patch has to be picked from *every* budget. Hence, the sum of all rows belonging to the same budget will give the row of ones. For example, in the simple-setup

(2 budgets per time period),

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

So the sum of the fist two rows is equal to the sum of the last two rows. However, Theorem 3 is more general than we need for the recursive characterization of DRUM as the following theorem demonstrates. Let B^t be the matrix from the $\mathcal{H}$-representation of a cone generated by A^t.

Theorem 4. *Assume that $\sum_{j_t \in \mathcal{J}^t}(I^t_{j_t} - 1) \leq |\mathcal{R}^t| - 1$ for all $t \in \mathcal{T}$. Then P is rationalized by DRUM if and only if (i) ρ is stable (ii) $\left(\otimes_{t \in \mathcal{T}} B^t\right) \rho \geq 0$.*

Condition $\sum_{j_t \in \mathcal{J}^t}(I^t_{j_t} - 1) \geq |\mathcal{R}^t| - 1$ imposes restrictions on the size of matrix A^t. It is satisfied in simple-setup and all the examples we are aware of. We conjecture this condition to be true always. Note that stability is a set of equality restrictions on ρ. Since any equality restriction can be represented as two inequality restrictions, Theorem 4 allows us to obtain the $\mathcal{H}$-representation of DRUM from the $\mathcal{H}$-representation of its static components recursively for any time window. In other words, one just needs to derive the $\mathcal{H}$-representation of RUM and then easily 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.

We highlight that 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

$$
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & - & - & - & - & - & - & - & - & - & - & - & - & - \\
- & - & - & - & - & - & - & - & - & - & - & - & 1 & 1 & 1 & 1 & 1 & - & - & - & - & - & - & - & - \\
- & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & - & 1 & 1 & 1 & 1 & 1 & - & - & - \\
- & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & - & - & - & - & - & - & - & - & 1 & 1 & 1 & 1 & - & 1 & 1 & - & - & - & 1 & 1 & - \\
- & - & - & - & 1 & 1 & 1 & 1 & - & - & - & - & - & - & - & - & - & - & - & 1 & - & - & - & - & - \\
- & - & - & - & - & - & - & - & 1 & 1 & - & - & - & - & - & - & 1 & - & - & - & 1 & - & - & - & 1 \\
- & - & - & - & - & - & - & - & - & - & 1 & 1 & - & - & - & - & - & - & - & - & - & 1 & - & - & - \\
1 & - & - & - & 1 & - & - & - & 1 & - & 1 & - & 1 & - & - & - & 1 & 1 & - & 1 & 1 & 1 & 1 & - & 1 \\
- & 1 & - & - & - & 1 & - & - & - & - & - & - & - & 1 & - & - & - & - & 1 & - & - & - & - & 1 & - \\
- & - & 1 & - & - & - & 1 & - & - & 1 & - & 1 & - & - & 1 & - & - & - & - & - & - & - & - & - & - \\
- & - & - & 1 & - & - & - & 1 & - & - & - & - & - & - & - & 1 & - & - & - & - & - & - & - & - & -
\end{pmatrix}
\begin{pmatrix}
x^t_{1|1} \\ x^t_{2|1} \\ x^t_{3|1} \\ x^t_{4|1} \\ x^t_{1|2} \\ x^t_{2|2} \\ x^t_{3|2} \\ x^t_{4|2} \\ x^t_{1|3} \\ x^t_{2|3} \\ x^t_{3|3} \\ x^t_{4|3}
\end{pmatrix}
$$

Table 5 – A^t for 3 goods and 3 budgets.

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. Nevertheless, our result implies that once the generic $\mathcal{H}$-representation of RUM becomes available, the analogous DRUM characterization will also become available.

5.3. Generalization of $\mathbb{D}$-monotonicity for 3 goods and 3 budgets per time period.

Now we showcase how our generalization of the Weyl-Minkowski theorem can be used to take the $\mathcal{H}$-representation of (static) RUM for the case of 3 goods and 3 budgets and use it to construct an $\mathcal{H}$-representation for DRUM for any time window. We consider a setup where each time period has 3 budgets with maximal intersections as in Example 3.2 in KS. The $\mathcal{V}$-representation in this case is given by matrix A_t in Table 5.

$x^t_{1\|1}$	$x^t_{2\|1}$	$x^t_{3\|1}$	$x^t_{4\|1}$	$x^t_{1\|2}$	$x^t_{2\|2}$	$x^t_{3\|2}$	$x^t_{4\|2}$	$x^t_{1\|3}$	$x^t_{2\|3}$	$x^t_{3\|3}$	$x^t_{4\|3}$
0	0	0	-1	0	0	0	-1	1	1	1	0
0	0	0	-1	1	0	0	0	1	0	0	0
1	0	0	0	1	0	0	0	0	0	0	-1
1	0	0	0	0	0	0	-1	1	0	0	0
1	0	1	0	-1	0	-1	0	0	0	0	0
0	-1	0	-1	1	0	1	0	0	0	0	0
0	0	0	0	1	1	0	0	0	-1	0	-1
0	0	-1	-1	0	0	0	0	1	1	0	0

Table 6 – The $\mathcal{H}$-representation of RUM for 3 goods and 3 budgets excluding nonnegativity and adding-up constraints.

The $\mathcal{H}$-representation in this case is displayed in Table ??6 (without the nonnegativity and adding-up constraints). As a consequence of the pattern of intersections, we have that $B^t = B^s$ for any $t, s \in \mathcal{T}$. Then we can establish the following result.

Corollary 2. *Let $K = 3$ and let $J^t = 3$ budgets for all $t \in \mathcal{T}$, such that the budgets have a* maximal pattern of intersections, *then the following are equivalent:*

(i) P with ρ vector representation is consistent with DRUM.

(ii) For $B^3 = \otimes_{t\in\mathcal{T}} B^3_t$, $B^3\rho \geq 0$ and ρ satisfies stability.

The proof of the previous result is direct from Theorem 4 since A^t in this case is such that $\sum_{j_t\in\mathcal{J}^t}(I^t_{j_t} - 1) = 10 \leq 24 = |\mathcal{R}^t| - 1$.

The case of 3 goods and 3 budgets with maximal intersection pattern in budgets is important because it shows what conditions D-monotonicity is missing. In particular, we can say that that the last 3 rows of the matrix displayed in Table 6 can be captured by D-monotonicity, as we will clarify this next. The rest of conditions are new.

Computationally Simple Testable Implications of the General Model.– Note that if $\mathrm{P} = (\mathrm{P}_{\mathbf{j}})_{\mathbf{j}\in\mathbf{J}}$ is consistent with DRUM, then the stochastic demand system consisting of 2 different budgets paths $(\mathrm{P}_{\mathbf{j}}, \mathrm{P}_{\mathbf{j}'})$ would also be consistent with DRUM. 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. Given $(\mathrm{P}_{\mathbf{j}}, \mathrm{P}_{\mathbf{j}'})$, at every t we can construct the implied four patches, which will be unions of patches constructed from P. And then check whether the implied distribution over these patches $\rho_{\mathbf{j},\mathbf{j}'}$ is D-monotone.

Proposition 1. If P is rationalized by DRUM, then (i) ρ is stable; (ii) for any two different budget paths $\mathbf{j}$ and $\mathbf{j}'$, the implied simple-setup $\rho_{\mathbf{j},\mathbf{j}'}$ is D-monotone.

Proof. The proof for necessity of stability follows from the proof of Theorem 2 since it does not use any features of the simple-setup. Necessity of D-monotonicity for any pair of budget paths follows from Theorem 2 and the fact that if P is consistent with DRUM, then any pair $(\mathrm{P}_{\mathbf{j}}, \mathrm{P}_{\mathbf{j}'})$ is consistent with DRUM. ■

Stability is applied to the distribution over all patches implied by P. That is, there is no need to use simple-setup. Stability and D-monotonicity are no longer sufficient in the general case. Theorem 1 has necessary and sufficient conditions for DRUM consistency but in some cases, it will be computationally more convenient to check the conditions in Proposition 1. We also highlight that in stark contrast with the static Weak Axiom of Stochastic Revealed Preference (Hoderlein and Stoye, 2014), our D-monotonicity condition has more empirical bite that is increasing with a longer time window $\mathcal{T}$. In that sense, with sufficiently long panels, we could obtain informative counterfactual bounds on the longitudinal distribution of demand out-of-sample, using only D-monotonicity. We remark that for the case of $K = 2$ goods and finitely many budgets per time period applying our results in Proposition 1 for all possible pairs of budgets paths is a necessary and sufficient test of DRUM.[16].

5.4. Characterization of DRUM via recursive Block Marschak inequalities

As we mentioned before a BM-like characterization of RUM in the demand setup does not exist. Here, we provide a characterization based on the BM inequalities while taking into account the limited menu variability in the demand setup and the monotonicity of utilities. We start with the observation that thanks to KS our problem can be discretized, so the main obstacle we overcome is to deal with the limited observability of choice sets formed by patches. Recall that $\mathcal{J}^t$ denotes the set of observed budgets at time t and every budget $j_t \in \mathcal{J}_t$ forms a collection of patches that belong to it. Thus, one can think of each budget as a menu of patches (i.e., consumers choose a patch from a menu $\{x^t_{i_t|j_t}\}_{i_t \in \mathcal{I}^t_{j_t}}$). Let $\mathbf{X}^t = \cup_{j \in \mathcal{J}^t}\{x^t_{i_t|j_t}\}_{i_t \in \mathcal{I}^t_{j_t}}$ be the "virtual" choice set–the set of all patches at time t. Also let $\bar{\mathcal{J}}^t = \{1, 2, \ldots, 2^{|\mathbf{X}^t|} - 1\}$ be the "virtual" set of budgets such that there is a one-to-one mapping between $j_t \in \bar{\mathcal{J}}^t$ and

[16]This is a consequence of Hoderlein and Stoye (2015) and Theorem 4

a nonempty subset of $\mathbf{X}^t$. We also assume that this mapping is such that the first $\mathcal{J}^t$ indexes correspond to observed budgets B_j^t. That is $\mathcal{J}^t$ is the set of all observed budgets and $\bar{\mathcal{J}}^t \setminus \mathcal{J}^t$ is the set of all "virtual budgets" that were not observed in the data. Using this extended definition of budgets, we can define a set of all (including the observed ones) budget paths $\bar{\mathbf{J}}$. Note that the set of observed budget paths $\mathbf{J}$ is a subset of $\bar{\mathbf{J}}$. Finally, as before, given a budget 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 vector representation 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 $\bar{\rho}$ that will be needed for our analysis.

Definition 8. We say that $\bar{\rho}$ *agrees* with ρ if they coincide on observed budget 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 $\mathbf{i} \in \mathbf{I_j}$.

This definition captures the idea of extension of ρ to budget paths that are not observed in the data (i.e. "virtual" budget paths).

Definition 9 (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 $t \in \mathcal{T}$ such that for every $y \in x_{i_t|j_t}$ there exists $i_t' \in \mathcal{I}_{j_t}^t$ and $y' \in x_{i_t'|j_t}$ such that $y' \geq y$.

IU-consistency captures the empirical content of strict monotonicity of the utility functions.

Given that each index j corresponds to a set of patches, let $C(j)$ denote the set of all patches in budget j. That is, $C(j_t) = \{x_{i_t|j_t}\}_{i_t\in\mathcal{I}_{j_t}^t}$.

Definition 10 (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 $\mathbf{i} \in \mathbf{I}_{\mathbf{j}}$

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

Note that the BM inequalities are linear in $\bar{\rho}$. Hence, we can construct matrix $\bar{B}^t$ with elements in $\{-1, 0, 1\}$ such that each row of $\bar{B}^t$ corresponds to some t, $\mathbf{j} \in \bar{\mathbf{J}}$, and $\mathbf{i} \in \mathbf{I}_{\mathbf{j}}$ and encodes one BM inequality.

With this definitions at hand we are ready to state the two main results of this subsection. First, a BM characterization of RUM for our demand setup and later an analogous characterization for DRUM.

Lemma 2. *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 $\bar{B}^t \bar{\rho} \geq 0$.

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 $\mu(\arg\max_{y \in B^t_{j_t}} u^t(y) \in x_{i_t|j_t}) = \rho(x_{i_t|j_t})$ 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 collection of patches j_t and $x_{i_t|j_t}$ in it such that $\bar{\rho}(x_{i_t|j_t}) > 0$ and for all $y \in x_{i_t|j_t}$ there exists a patch $x_{i'_t|j_t} \in C(j_t)$ and $y' \in x_{i'_t|j_t}$ such that $y' \geq y$ and $y' \neq y$. But this is not possible since u^t is assumed to be a monotone function (i.e, $u^t(y) < u^t(y')$), 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 $\mathbf{X}^t$. By the result in Falmagne (1978),

we know that there is a $\nu \in \Delta\left(\bar{\mathcal{R}}^t\right)$ such that

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

Since ρ is IU-consistent, $\nu(\succ) = 0$ for any $\succ \in \bar{\mathcal{R}}^t$ that is not an extension of the strict vector order $>$. To see this is true, we will prove the contrapositive. Namely, if $\nu(\succ) > 0$ for some $\succ \in \bar{\mathcal{R}}^t$ that is not an extension of the strict vector order $>$, in particular, there exists $y, x \in \mathbf{X}^t$ such that $x > y$ yet $y \succ x$, then IU-consistency fails for the virtual budget $\{y, x\}$.

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

In sum, since the BM inequalities with the nonnegativity and adding-up constraints for all menus provide the $\mathcal{H}$-representation of RUM for the static case and as a direct application of Theorem 3, we can establish the following result.

Theorem 5. *The following are equivalent.*

(i) P with a vector representation ρ is consistent with DRUM.

(ii) There exists $\bar{\rho}$ that agrees with ρ, is IU-consistent, stable, and satisfies $(\otimes_{t\in\mathcal{T}}\bar{B}^t)\bar{\rho} \geq 0$.

Proof. (i) implies (ii). Direct from arguments analogous to those made in Lemma 2, Theorem 3, and Theorem 4.

(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 $(\otimes_{t\in\mathcal{T}}\bar{B}^t)\bar{\rho} \geq 0$ and stability, we can use the results in Theorem 3 and the results in Theorem 4, to ensure that there exists a $\nu \in \Delta(\bar{\mathcal{R}})$ such that

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

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 strict vector order $>$. To see this is true, we will 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 strict vector order $>$, in particular, there exist $y,x\in\mathbf{X}^t$ such that $x>y$ yet $y\succ x$, then IU-consistency fails for the virtual budget path that contains the budget $\{y,x\}$ at time t. ■

Theorem 5 is not exactly a $\mathcal{H}$-representation of DRUM. It becomes one when all menus in $\bar{\mathbf{J}}$ are observed like it is assumed in Chambers *et al.* (2021), Li (2021). Note, moreover, that our proof can be used in the environments in Chambers *et al.* (2021) and Li (2021) without changes. The vector order could be replaced by any primitive order, including the empty order such as the abstract setup in those two papers. We generalize the BM inequalities for the case of unobserved menus. Even if for our primitive this recursive characterization of DRUM is not an $\mathcal{H}$-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; (ii) it provides additional intuition about the additional empirical bite of DRUM in comparison with RUM; and (iii) it means DRUM can be tested with a linear program.

To fully 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 11 (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

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

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

Now we establish the following result.

Corollary 3. *If P with a vector representation ρ is consistent with DRUM then ρ satisfies the DRUM-BM inequalities.*

This result shows that DRUM has a rich set of new constraints beyond what BM inequalities require in the static case.

6. Endogenous Expenditure with Dynamic Random Augmented Utility Model–DRAUM

So far, we have assumed that budgets are exogenously given. In particular, the definition of DRUM requires the probability measure over utility functions to be independent of budgets (prices and income). This assumption is satisfied in experimental setups such as the one in Porter and Adams (2016). But this assumption may not be realistic in other setups, mainly, when saving is possible. In this section, we relax the exogeneity assumption by extending the results in Deb *et al.* (2021) to our setup. Our new model will cover the classical consumption smoothing problem with income uncertainty.

Similarly to DRUM we can define a Dynamic Random Augmented Utility Model (DRAUM). Let V denote the set of all continuous, strictly concave, and monotone augmented utility functions that map $X^* = X \times \mathbb{R}_-$ to $\mathbb{R}$ and $\mathcal{V} = \times_{t\in\mathcal{T}} V$ be the Cartesian product of T repetitions of V.

Definition 12 (DRAUM)**.** A dynamic stochastic demand P is consistent with DRAUM 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 DRUM is an extension of RUM to dynamic settings (i.e., DRUM and RUM coincide when $T = 1$), DRAUM is a dynamic extension of the Random Augmented Utility Model (RAUM) of Deb *et al.* (2021). For $T \geq 2$, the DRAUM does not restrict the dependence of the augmented utility v^t across time and allows full heterogeneity in cross-sections.

Example continued [Consumption Smoothing with Income Uncertainty Example 3 continued]

Recall 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 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 in this case, the state-dependent 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 DRAUM.

Example 3 illustrates that DRAUM covers, as a particular case, the critical case of consumption smoothing with income uncertainty. Note that the correlation among consumption in time induced by consumption smoothing is rich but covered by DRAUM. In this setup, the random augmented utility stochastic process is independent of prices because prices are determined exogenously by supply and demand forces.

Next, we characterize DRAUM by using the fact that consistency with DRAUM is equivalent to consistency with DRUM 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 (i.e., $B_{\hat{j}^t} = \{y \in X : p_{\hat{j},t}'y = 1\}$). Using these normalized budgets, we can define patches in complete analogy to the general case, and then we can obtain

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

for all $\mathbf{i} \in \mathbf{I_j}$, $\mathbf{j} \in \mathbf{J}$. We define the projected vector representation of P as

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

Similarly to the simple-setup, we rule out the intersection patches. That is, we assume that for any $\mathbf{j}$ and $\mathbf{j}'$

$$\mathrm{P}_{\mathbf{j}}\left(\left\{y^t \in X \ : \ y^t/p'_{j_t,t}y^t \in B^t_{j_t} \quad \text{and} \quad y^t/p'_{j'_t,t}y^t \in B^t_{j'_t}\right\}_{t\in\mathcal{T}}\right) = 0$$

whenever $j_t \neq j'_t$ for some t.

Lemma 3. *The following are equivalent:*

(i) The dynamic stochastic demand P is consistent with DRAUM.

(ii) There exists $\nu \in \mathbb{R}^{|\mathcal{R}|}_+$ such that $\rho_* = A\nu$.

The proof is omitted because it is analogous to our main Theorem 1, using the same logic as in Theorem 3 in Deb *et al.* (2021), which shows that in static settings checking consistency with Random Augmented Utility Model is equivalent to checking consistency with RUM in the projected stochastic demand with budgets with income equal to 1. Since DRAUM requires rationalizability by an augmented utility in each period, our result follows immediately.

7. Relationship with Afriat's and McFadden-Richter's frameworks

Using Theorem 1, we proceed to study the implications of DRUM for simpler domains than a panel dataset. In particular, first we look at a time series with the assumption of constant utility across time periods as in Afriat's framework. In this case, DRUM implies that the (deterministic) Strong Axiom of Revealed Preference (SARP) has to hold in time series. Next 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 corresponds to one or many time periods, respectively. We show that if P 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. This means that DRUM has more empirical bite than RUM in our domain.

7.1. Afriat's framework

DRUM has no testable implications for a time series without further restrictions. That is, if we observe $\mathrm{P}_{\mathbf{j}}$ 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 Afriat's framework, one only needs time-series of choices from budgets to test utility maximization. The reason for this is that in 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.

Definition 13 (Strong Axiom of Revealed Path Dominance, SARPD)**.** For a given $\mathbf{j} \in \mathbf{J}$,

$\rho_{\mathbf{j}} = (\rho(x_{\mathbf{i}|\mathbf{j}}))_{\mathbf{i}\in\mathbf{I_j}}$ 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 $x^{t}_{i_t|j_t} \succeq^* x^{s}_{i_s|j_s}$ if and only if for some $x \in x^{t}_{i_t|j_t}$ 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 Afriat's framework. Using SARPD, we can establish the following result.

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

We provide here the proof of Proposition 2 because of its simplicity and interest. Assume towards contradiction that P 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 violations 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. In simple words, SARPD rules out the possibility that there are some individuals in the population that violate SARP. Yet again, constancy of DRUM is what drives testability in a single budget path or time series. When constancy is relaxed, we need to obtain cross-sectional variation (i.e., more than one budget path) to reestablish testability of DRUM.

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. Evidently, $\mathbb{D}$-monotonicity is not exhausted with the previous inequality, but DRUM restricts the probability of choice paths with RP cycles meaningfully. This *endogenous* bounds on the probability of a choice path that contains a revealed preference cycle has an important advantage with respect to measures of deviations to rationality like the Critical Cost Efficiency Index (CCEI) (Afriat, 1967), because in that literature it is very hard to set a threshold on what is the level of deviations from static utility maximization that is deemed reasonable. In our setup, we convert this problem into a population one and then just 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 $\mathbb{D}$-monotonicity is equivalent to the Weak Axiom of Revealed Preference by Samuelson (1938) in 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.

7.2. 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 given 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 data. The slicing distribution is a mixture of marginal distributions. It corresponds to the situations when the researcher only focuses on one cross-section.

Proposition 3. If P 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 $\mathbf{j}\in\mathbf{J}$.

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 4. *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. ∎

Proposition 3 provides a set of necessary conditions that can substantially simplify the testing. One just need to jointly test all possible marginal and conditional distributions for being RUM consistent. Moreover, Proposition 3 means that if P is consistent with DRUM then the data in any given cross-section (slice) is consistent with RUM. 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.

Example 5. [Marginals are consistent with WASRP but not rationalized by DRUM] Consider ρ presented in Table 7. 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.[17] 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.

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

	$x^2_{1\|1}$	$x^2_{2\|1}$	$x^2_{1\|2}$	$x^2_{2\|2}$
$x^1_{1\|1}$	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 7 – Matrix representation of ρ that is consistent with RUM after slicing, but is not consistent with DRUM

7.3. 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 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 $B^t_j \neq B^{t'}_{j'}$ 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 14 (Pooled Patches)**.** Let

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

be the coarsest partition of $\bigcup_{t\in\mathcal{T}} \bigcup_{j\in\mathcal{J}^t} B^t_j$ such that

$$\xi^t_{k|j} \bigcap B^t_{j'} \in \{\xi^t_{k|j}, \emptyset\}$$

for any j, j' and k.

The pooled patches $\{\xi^t_{k|j}\}$ partition every $x^t_{i|j}$ since B^t_j now may intersect with budgets from different from t time periods (see Figure 3). Given these new patches, we can define the pooled

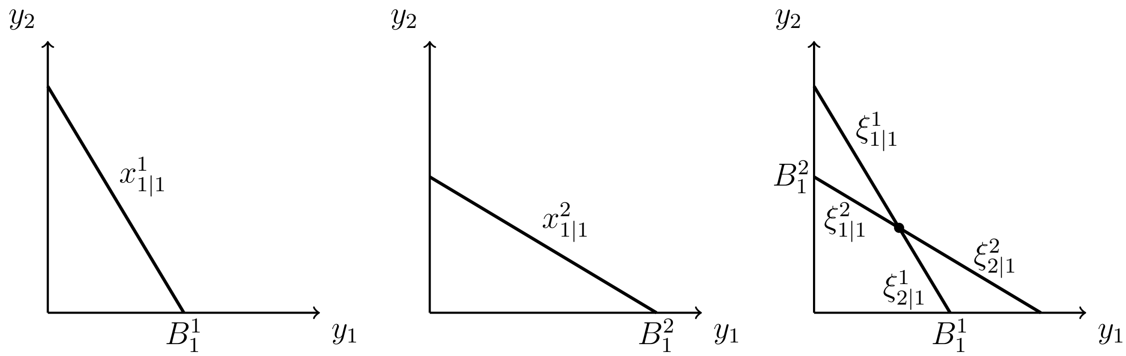

Figure 3 – $K = 2$ goods, $T = 2$ time periods, one budget per time period. The first and the second picture depict patches in 2 time periods. The third picture depicts new patches that arise after pooling the data.

demand $\rho^{\text{pool}}\left(\xi^t_{k|j}\right)$ as the probability of observing someone picking from patch $\xi^t_{k|j}$. Next we construct a simple 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^t_1. Assume that $B^1_1 \neq B^2_1$ and $B^1_1 \cup B^2_1 \neq \emptyset$ (see Figure 3). Given that there is no budget variation for any given time period, there is only one choice path $\left(x^1_{1|1}, x^2_{1|1}\right)$. So the trivial $\rho\left(\left(x^1_{1|1}, x^2_{1|1}\right)\right) = 1$ is rationalizable by DRUM. After pooling, since the budgets overlap, there are 4 patches (we assume that the demand is continuous so there is no intersection patch). Since there is only one choice path, DRUM does not impose any restrictions on choice of individuals in these two budgets. As a result, we can take ν^1 and ν^2 from the DRUM definition such that $\rho^{\text{pool}}\left(\xi^2_{1|1}\right) + \rho^{\text{pool}}\left(\xi^1_{2|1}\right) > 1$. As a result, this ρ^{pool} cannot be consistent with RUM.

8. Counterfactuals

This section shows how one can conduct sharp counterfactual analyses within our framework.[18] The sharpness of our results follows from the fact that we have a full characterization of DRUM.[19] To simplify the exposition, we will focus on the simple-setup. That is, we will consider settings with two intersecting budgets per time period.

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 at $T+1$ the consumers face a pair of prices $p_{1,T+1}$ and $p_{2,T+1}$ that are known to the analyst, let the income at both budgets be 1 (recall that we are working with the simple-setup). We will 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}.

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 15 (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

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

[18]Sharpness in this setting means that we can compute the shortest possible sets of parameters that are consistent with the observed data and the model.

[19]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.

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

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

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 $g: X\rightarrow\mathbb{R}$, let

$$\underline{g}\left(x_{i_t|j_t}^{t}\right)=\inf_{y\in x_{i_t|j_t}^{t}}g(y),$$
$$\overline{g}\left(x_{i_t|j_t}^{t}\right)=\sup_{y\in x_{i_t|j_t}^{t}}g(y).$$

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

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

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

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

Note that our results are complementary with Kitamura and Stoye (2019) that predicts 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.

9. Monte Carlo Simulations: Statistical Test of DRUM

This section provides a Monte Carlo study to evaluate the performance of the KS's test in testing DRUM in finite samples. We refer to KS for a formal definition of the testing procedure. We consider the simple-setup used throughout the paper with $K = T = J^t = 2$. We set the number of consumers per choice path to $N \in \{50, 500, 5000\}$ and the number of simulations for each data generating process (DGP) to $R = 1000$. The critical values for each test statistic are computed using $B = 999$ bootstrap samples. The tuning parameter τ from KS is set to $\tau = \sqrt{\log(4N)/4N}$ as recommended in KS (given that there are 4 choice paths in every budget path $4N$ 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 time periods are the same and correspond to prices $(2, 1)'$ and $(1, 2)'$ and expenditure level 1.

We consider 2 DGPs for random $\alpha = (\alpha_1, \alpha_2)'$.

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

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 5% significance level for all 3 sample sizes and both DGPs are presented in Table 8 The rejection rates are close to 5% even for small sample sizes.

DGP	N	Rejection rate, %
DGP1	50	3.4
	500	4.3
	5000	5.1
DGP2	50	3.7
	500	4.6
	5000	5.4

Table 8 – Every entry represents the rejection rate at 5% significance level and is computed from 1000 simulations and 999 bootstraps per simulation.

To analyze the finite sample power of the test, we consider the DGP from Table 7. 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 7 has marginal probabilities consistent with RUM. Yet, even at small sample sizes such as $N = 50$ rejection rate is 100%. This simulations show that the KS test of DRUM has good size and power properties in finite samples.

10. Conclusion

We have introduced and 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 consumers keep their preferences stable over time. This generality is essential because the static utility maximization framework often fails to explain 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 Afriat's and McFadden-Richter's frameworks combining their strengths and reducing their weaknesses.

We have obtained 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. *International economic review*, **8** (1), 67–77. (document), 1, 7.1

Aguiar, V. H. and Kashaev, N. (2021). Stochastic revealed preferences with measurement error. *The Review of Economic Studies*, **88** (4), 2042–2093. 1, 6, 2, 19

—, — and Allen, R. (2022). Prices, profits, proxies, and production. *Journal of Econometrics*. 19

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

Allen, R. and Rehbeck, J. (2019). Identification with additively separable heterogeneity. *Econometrica*, **87** (3), 1021–1054. 19

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

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, 12

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, 5.2

Blundell, R., Browning, M. and Crawford, I. (2008). Best nonparametric bounds on demand responses. *Econometrica*, **76** (6), 1227–1262. 19

—, Kristensen, D. and Matzkin, R. (2014). Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics*, **179** (2), 112–127. 19

Border, K. (2007). Introductory notes on stochastic rationality. *California Institute of Technology*. 3.1, 11.1

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

Cattaneo, M. D., Ma, X., Masatlioglu, Y. and Suleymanov, E. (2020). A random attention model. *Journal of Political Economy*, **128** (7), 2796–2836. 1

Chambers, C. P., Masatlioglu, Y. and Turansick, C. (2021). Correlated choice. *arXiv preprint arXiv:2103.05084*. 1, 14, 5.4

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). â€˜regularâ€™choice and the weak axiom of stochastic revealed preference. *Economic Theory*, **31** (1), 35–50. 4.1

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*. 4, 5, 1, 2, 8, 6, 6, 6, 7.3

Dogan, S. and Yildiz, K. (2022). Every choice function is pro-con rationalizable. *Operations Research*. 11.4

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. 5.2, 5.4

Frick, M., Iijima, R. and Strzalecki, T. (2019). Dynamic random utility. *Econometrica*, **87** (6), 1941–2002. 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, 4.1, 5.2, 5.3

— and — (2015). Testing stochastic rationality and predicting stochastic demand: the case of two goods. *Economic Theory Bulletin*, **3** (2), 313–328. 13, 16

Im, C. and Rehbeck, J. (2021). Non-rationalizable individuals, stochastic rationalizability, and sampling. *Available at SSRN 3767994*. 1, 10

Kawaguchi, K. (2017). Testing rationality without restricting heterogeneity. *Journal of Econometrics*, **197** (1), 153–171. 1, 3.1, 3.1

Kitamura, Y. and Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, **86** (6), 1883–1909. 1, 5, 5.1, 5.2, 7.3

— and — (2019). Nonparametric counterfactuals in random utility models. *arXiv preprint arXiv:1902.08350*. 8

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

Li, R. (2021). An axiomatization of stochastic utility. *arXiv preprint arXiv:2102.00143*. 1, 14, 5.4

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

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, 3.1, 7, 7.2

McFadden, D. L. (2005). Revealed stochastic preference: a synthesis. *Economic Theory*, **26** (2), 245–264. 3.1, 7

Norets, A. and Tang, X. (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies*, **81** (3), 1229–1262. 19

Polisson, M., Quah, J. K.-H. and Renou, L. (2020). Revealed preferences over risk and uncertainty. *American Economic Review*, **110** (6), 1782–1820. 1

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, 6

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. 11.4

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica*, **5** (17), 61–71. 1, 7.1

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. 3.1, 3.1, 5.1

Stoye, J. (2019). Revealed stochastic preference: A one-paragraph proof and generalization. *Economics Letters*, **177**, 66–68. 5.2

Strzalecki, T. (2021). *Stochastic Choice*. Mimeo. 1, 4.2

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica: Journal of the Econometric Society*, pp. 945–973. 19

— (1984). The nonparametric approach to production analysis. *Econometrica: Journal of the Econometric Society*, pp. 579–597. 19

11. Proofs

11.1. Proof of Theorem 1

$((i) \iff (ii) \iff (iii))$

In this proof, we adapt the proof of Theorem 3.1 in KS for RUM for the dynamic case. Our proof uses profiles of nonstochastic demand profiles. For each time period $t \in \mathcal{T}$ we define the 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 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 representation corresponding to columns of A_T, the statement of the theorem follows immediately.

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

11.2. Proof of Theorem 2

Necessity. Suppose that ρ is rationalized by DRUM.

Necessity of stability. By the definition of DRUM, there exists a distribution over $\mathcal{U}$, μ, such that

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

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

$$\sum_{i\in\mathcal{I}_{j_{t'}}^{t'}} \rho\left(x_{\mathbf{i}|\mathbf{j}}\right) =$$

$$\sum_{i\in\mathcal{I}_{j_{t'}}^{t'}} \int \mathbb{1}\left(\arg\max_{y\in B_{j_{t'}}^{t'}} u^{t'}(y) \in x_{i|j_{t'}}^{t'}\right) \prod_{t\in\mathcal{T}\setminus\{t'\}} \mathbb{1}\left(\arg\max_{y\in B_{j_t}^t} u^t(y) \in x_{i_t|j_t}^t\right) d\mu(u) =$$

$$\int \sum_{i\in\mathcal{I}_{j_{t'}}^{t'}} \mathbb{1}\left(\arg\max_{y\in B_{j_{t'}}^{t'}} u^{t'}(y) \in x_{i|j_{t'}}^{t'}\right) \prod_{t\in\mathcal{T}\setminus\{t'\}} \mathbb{1}\left(\arg\max_{y\in B_{j_t}^t} u^t(y) \in x_{i_t|j_t}^t\right) d\mu(u) =$$

$$\int \prod_{t\in\mathcal{T}\setminus\{t'\}} \mathbb{1}\left(\arg\max_{y\in B_{j_t}^t} u^t(y) \in x_{i_t|j_t}^t\right) d\mu(u),$$

where the last equality follows from $\arg\max_{y\in B^{t'}_{j_{t'}}} u^{t'}(y)$ being a singleton ($u^{t'}$ is continuous and strictly monotone) 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. Note that stability is necessary for DRUM to hold not just in simple-setup but in settings with many budgets.

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 the 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^1_2, and ν^1_3) and ρ into 4 vectors ($\rho^1_{1|1}$, $\rho^1_{2|1}$, $\rho^1_{1|2}$, and $\rho^1_{2|2}$) such that

$$\begin{pmatrix} \rho^1_{1|1} \\ \rho^1_{2|1} \\ \rho^1_{1|2} \\ \rho^1_{2|2} \end{pmatrix} = \rho = A_T\nu = A_1 \otimes A_{T-1}\nu = \begin{pmatrix} A_{T-1} & A_{T-1} & 0 \\ 0 & 0 & A_{T-1} \\ A_{T-1} & 0 & 0 \\ 0 & A_{T-1} & A_{T-1} \end{pmatrix} \begin{pmatrix} \nu^1_1 \\ \nu^1_2 \\ \nu^1_3 \end{pmatrix} = \begin{pmatrix} A_{T-1}(\nu^1_1 + \nu^1_2) \\ A_{T-1}\nu^1_3 \\ A_{T-1}\nu^1_1 \\ A_{T-1}(\nu^1_2 + \nu^1_3) \end{pmatrix}.$$

In this representation, $\rho^1_{i|j}$ correspond to all choice paths that contain patch $x^1_{i|j}$. Subtracting the third line from the first one, and the second line from the forth 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 $x^1_{i'_1|j'_1} \succ^D x^1_{i_1|j_1}$.

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 $\rho^1_{i|j,i'|j'}$ correspond 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 $x^1_{i'_1|j'_1} \succ^D x^1_{i_1|j_1}$ and $x^2_{i'_2|j'_2} \succ^D x^2_{i_2|j_2}$. 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 $x^t_{i'_t|j'_t} \succ^D x^t_{i_t|j_t}$ 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 $B_L = B^1 \otimes B_{L-1}$ and $P_{A_T} = P_{A_1} \otimes P_{A_{T-1}}$, where

$$B_1 = (A'_1A_1)^{-1}A'_1 = \begin{pmatrix} 0.25 & 0.25 & 0.75 & -0.25 \\ 0.5 & -0.5 & -0.5 & 0.5 \\ -0.25 & 0.75 & 0.25 & 0.25 \end{pmatrix}$$

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 $\nu = B_T\rho$ 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

$$\begin{pmatrix} \nu_1^1 \\ \nu_2^1 \\ \nu_3^1 \end{pmatrix} = \nu = B_T\rho = B_1 \otimes B_{T-1}\rho = \begin{pmatrix} 0.25B_{T-1} & 0.25B_{T-1} & 0.75B_{T-1} & -0.25B_{T-1} \\ 0.5B_{T-1} & -0.5B_{T-1} & -0.5B_{T-1} & 0.5B_{T-1} \\ -0.25B_{T-1} & 0.75B_{T-1} & 0.25B_{T-1} & 0.25B_{T-1} \end{pmatrix} \begin{pmatrix} \rho_{1|1}^1 \\ \rho_{2|1}^1 \\ \rho_{1|2}^1 \\ \rho_{2|2}^1 \end{pmatrix}.$$

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

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

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

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

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

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

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 $x^t_{i'_t|j'_t} \succ^D x^t_{i_t|j_t}$ 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, $A\nu = \rho$.

11.3. Proof of Theorem 3

Let $L_T = \otimes_{t\in\mathcal{T}} L^t$ and $K_T = \otimes_{t\in\mathcal{T}} K^t$.

Step 1. Take any $v \geq 0$. We want to show that $L_T K_T v \geq 0$. Note that by the properties of the Kronecker product

$$L_T K_T v = \otimes_{t=1}^{T}(L^t K^t)v.$$

Define $D^{T-1} = \otimes_{t=1}^{T-1}(L^t K^t)$. We can rewrite $L_T K_T v$ as

$$L_T K_T v = \begin{pmatrix} D^{T-1}_{11} L^T K^T & D^{T-1}_{12} L^T K^T & \dots \\ D^{T-1}_{21} L^T K^T & D^{T-1}_{22} L^T K^T & \dots \\ \dots & \dots & \dots \end{pmatrix} v = \begin{pmatrix} \sum_k D^{T-1}_{1k} L^T K^T v^k_T \\ \sum_k D^{T-1}_{2k} L^T K^T v^k_T \\ \dots \end{pmatrix},$$

where D^{T-1}_{ij} is the (i, j)-element of D^{T-1} and v^k_T are subvectors of v of the length equal to the number of columns of $L^T K^T$. By assumptions of the theorem, $L^t K^t \tilde{v} \geq 0$ for any $\tilde{v} \geq 0$

and any t, hence,

$$v \geq 0 \implies v_T^k \geq 0 \implies L^T K^T v_T^k \geq 0.$$

Moreover, the rows of $L_T K_T v$ are weighted sums of the elements of the rows of matrix D^{T-1} weighted with some non-negative weights $L^T K^T v_T^k$. Hence, if $D_{T-1} y \geq 0$ for any conformable vector $y \geq 0$, then $L_T K_T v \geq 0$. Repeating the above step one more time, we can conclude that $D_{T-1} y \geq 0$ for any $y \geq 0$, if $D_{T-2} x \geq 0$ for any conformable $x \geq 0$. Repeating the step finitely many times we obtain that if $D^1 = L^1 K^1 z \geq 0$ for any $z \geq 0$, then $L_T K_T v \geq 0$ for any $v \geq 0$. The latter is satisfied by the definition of L_1 and K_1. Hence,

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

Step 2. K_T has a full row rank if and only if K^t is of full row rank for all t. Thus, if K_T has full row rank then we can represent any z as a weighted sum of columns of K_T (some weights may be negative). Thus, we want to show that if $\{v \,:\, K_T v = z\}$ does not contain nonnegative vectors, then $L_T z$ should have a negative component. We will prove the result by induction. For $T = 2$, towards a contradiction, assume that $L_T z \geq 0$ and every $v \in \{v \,:\, K_T v = z\}$ has at least one negative component. Take any $v \in \{v \,:\, K_T v = z\}$. Since $L_T z = L_T K_T v \geq 0$, we can conclude that

$$L^T K^T \sum_k D_{mk}^{T-1} v_T^k \geq 0$$

for all m. Hence, by the definition of L^T, $\sum_k D_{mk}^{T-1} v_T^k \geq 0$ for all m. Let V_T be the matrix which k-th column is v_T^k. Hence,

$$D^{T-1} V_T' \geq 0$$

For $T = 2$, $D^{T-1} = L^1 K^1$. Thus, $D^{T-1} V_T' \geq 0$ implies that every element of V_T must be nonnegative. The later is not possible since $\{v \,:\, K_T v = z\}$ does not contain nonnegative vectors. Now assume that for $K-1$, the statement is correct. That is,

$$D^{K-1}v \geq 0 \iff v \geq 0.$$

Towards a contradiction assume that the statement is incorrect for $T = K$. Repeating the above argument, we will obtain that

$$D^{K-1}V'_K \geq 0.$$

Thus, $v \geq 0$. The contradiction completes the proof.

11.4. Proof of Theorem 4

Step 1. If ρ is consistent with DRUM, then stability is implied by the proof of Theorem 2. Moreover, $\otimes_{t\in\mathcal{T}} B^t\rho \geq 0$ follows from applying Theorem 3 to $K^t = A^t$ and $L^t = B^t$.

Step 2. Assume that ρ is stable and $\otimes_{t\in\mathcal{T}} B^t\rho \geq 0$. If we show that the system of equations $A_{\mathcal{T}}v = \rho$ has a solution, then Step 2 of the proof of Theorem 3 would imply that the system of equations $A_{\mathcal{T}}v = \rho$ has a nonnegative solution. Hence, ρ is consistent with DRUM.

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

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

First, construct matrix A^{t*} by removing the last row from A^t. That is, from every budget except the first one (the first two rows), we removed the rows that correspond 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 rows to 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,

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

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^*_T, as a Kronecker product of full row rank matrices, is of full row rank as well. Thus, $v^* = A^{*\prime}_T(A^*_T A^{*\prime}_T)^{-1}\rho^*$ exists and solves $A^*_T v = \rho^*$. If we show that $A^-_T v^* = \rho^-$, then v^* solves $A_T v = \rho$ 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_T and A^*_T) are defined as a function of $\mathcal{T}$. Hence, if we push the time period t to the very end (i.e., $1, \ldots, t-1, t+1, \ldots, T, t$), 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, $W^1\left[\rho^*\right]$ changes labels so that $t=1$ is the last one and reshuffles elements of ρ^*, then $diag(G^1)W^1\left[\rho^*\right]$ 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 $x^1_{2|2}$ dropped.

Let Y^t be an operator such that $Y^t[\cdot] = W^{t,-1}\left[diag(G^t)W^t[\cdot]\right]$. 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 budget, except the first one, we pick the last patch 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 patches removed from A^t. Consider the first row of A^{t-}. It corresponds to a last patch from the second budget at time t. Note that since sum of all rows that correspond to the same budget is equal to the row of ones. Hence, the first row of A^{t-} is equal to sum of the rows that correspond to budget 1 minus sum of the remaining rows in budget 2. That is, the first row of A^{t-} can be written as

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

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

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

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

Next note that, up to 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 the patches 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 patches. As a result, we can split the original system into two: $A_T^* v = \rho^*$ and $A_T^- v = \rho^-$.

Consider the system $A_T^* v = \rho^*$. 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, $A_T^* A^{*\prime}$ is invertible and $v^* = A^{*\prime}\left(A_T^* A^{*\prime}\right)^{-1} \rho^*$ solves the system. If we show that

$$A_T^- v^* = \rho^-,$$

then we prove that $A_T v = \rho$ 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 correspond to choice paths that contain one of the removed patches 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 correspond to choice path where patch 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 patches 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 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 5.4.

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

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

be the vector of 0s and 1s that the best patch in every virtual budget. Analogously to A^t, let $\bar{A}^t$ denote the matrix which columns are $\{a^t_{\succ}\}_{\succ\in\bar{\mathcal{R}}^t}$, 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 5. *$\bar{A}^{t*}$ 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 and any negative vector such that the sum over all patches in each budget does not depend on a budget. Moreover, since any vector can be written as a sum of a positive and a negative vectors, $\bar{A}^t\nu$ 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 $\bar{A}^{t*}\nu = \bar{\rho}^*$, where $\bar{\rho}^*$ is any vector. Hence, $\bar{A}^{t*}$ is of full row rank. Indeed, if it was not, then there would exists $\xi \neq 0$ such that $\xi'\bar{A}^{t*}\nu = 0 \cdot \nu = 0$ for all ν. Hence, $\xi'\bar{A}^{t*}\nu = \xi'\bar{\rho}^* = 0$ for all $\bar{\rho}^*$. The latter contradicts to $\xi \neq 0$. ■

Lemma 5 implies that A^{t*} is of full row rank since by assumption it has at least the same number of columns as rows ($\sum_{j_t\in\mathcal{J}^t}(I^t_{j_t} - 1) \leq |\mathcal{R}^t| - 1$) and can be constructed from $\bar{A}^t\nu$ by removing some rows and some columns.

We conclude this section by by the following proposition that can be used to simplify the computational complexity of the $\mathcal{H}$-representation.

Proposition 5. Assume that $\sum_{j_t\in\mathcal{J}^t}(I^t_{j_t} - 1) \leq |\mathcal{R}^t| - 1$. Then ρ is consistent with DRUM if and only if (i) ρ is stable and $\otimes_{t\in\mathcal{T}}B^{t*}\rho^* \geq 0$, where B^{t*} is the $\mathcal{H}$-representation of A^{t*}.

Proof. Step 1. Note that since $A^*_{\mathcal{T}}$ is a submatrix of $A_{\mathcal{T}}$, $A_{\mathcal{T}}v = \rho$ implies $A^*_{\mathcal{T}}v = \rho^*$. The latter implies that $\otimes_{t\in\mathcal{T}}B^{t*}\rho^* \geq 0$. Stability of ρ follows from Theorem 4.

Step 2. Assume that ρ is stable and that $\otimes_{t\in\mathcal{T}} B^{t*}\rho^* \geq 0$. By Theorem 3, since A_T^* has full row rank, there is $\bar{v} \geq 0$ such that $A_T^*\bar{v} = \rho^*$. It is left to show that $A_T^-\bar{v} = \rho^-$. Note that $\bar{v} = v^* + z$, where $v^* = A_T^{*\prime}(A_T^* A_T^{*\prime})^{-1}$ and $z = \bar{v} - v^*$. Since v^* is also a solution of $A_T^* v = \rho^*$, $A_T^* z = 0$. Since stability implies that $A_T^- v^* = \rho^-$ (see the proof of Theorem 4), it suffices to show that $A_T^- z = 0$. The latter follows from

$$\otimes_{t\in\mathcal{T}} C^t z = \circ_{t':C^{t'}=A^{t'-}} Y^{t'}\left[A_T^* z\right] = \circ_{t':C^{t'}=A^{t'-}} Y^{t'}\left[0\right] = 0.$$

■