Experimental and Clinical Psychopharmacology

Identification and Management of Nonsystematic Cross-Commodity Data: Toward Best Practice

Mark J. Rzeszutek, Sean D. Regnier, Brent A. Kaplan, Haily K. Traxler, Jeffrey S. Stein, Devin C. Tomlinson, and Mikhail N. Koffarnus

Online First Publication, March 5, 2026. https://dx.doi.org/10.1037/pha0000836

Experimental and Clinical Psychopharmacology

ISSN: 1064-1297

https://doi.org/10.1037/pha0000836

Identification and Management of Nonsystematic Cross-Commodity Data: Toward Best Practice

Mark J. Rzeszutek[1], Sean D. Regnier[2], Brent A. Kaplan[3], Haily K. Traxler[4], Jeffrey S. Stein[5], Devin C. Tomlinson[6], and Mikhail N. Koffarnus[1]

[1] Department of Family and Community Medicine, College of Medicine, University of Kentucky
[2] Department of Behavioral Science, College of Medicine, University of Kentucky
[3] codedbx, Lexington, Kentucky, United States
[4] Clinical Behavior Analysis, Louisville, Kentucky, United States
[5] Fralin Biomedical Research Institute, Virginia Polytechnic Institute and State University
[6] Department of Psychiatry, University of Michigan

Data systematicity has been an important area of consideration for behavioral economic demand. Stein et al. (2015) introduced criteria and an accompanying algorithm to aid researchers in identifying data series that may be considered "nonsystematic"—that is, data that may not follow empirically based assumptions such as an overall decrease in consumption as the cost of a commodity increases and consistency in decreases in consumption. However, those criteria and algorithm are only directly applicable to own-price demand, or demand for a commodity that is increasing in price. Cross-price demand, or demand for a second commodity that changes as a function of some other commodity, does not have a similar set of criteria or algorithm for assessing cross-commodity demand systematicity. Cross-price or cross-commodity demand is useful in understanding how changes in one substance or commodity may change the consumption of another substance or commodity. Thus, we extend Stein et al.'s criteria and algorithm to classify if a cross-commodity can be considered a substitute, complement, or independent, and then assess its systematicity based on its classification. We demonstrate this algorithm on three different cross-commodity demand data sets and describe important considerations regarding data exclusions to prevent biasing results from own-price and cross-price demand.

Public Health Significance

Cross-commodity demand is a useful method to assess how changes in consumption of one commodity affect consumption of another. However, there are currently no guidelines to assess nonsystematic cross-commodity demand data. We introduce methods and recommendations for assessing nonsystematicity for cross-commodity demand, which can help to identify procedural differences that affect cross-commodity demand systematicity.

Keywords: purchase task, data systematicity, behavioral economics, demand, sensitivity analysis

Supplemental materials: https://doi.org/10.1037/pha0000836.supp

Kelly E. Dunn served as action editor.

Mark J. Rzeszutek https://orcid.org/0000-0003-4555-9767

Data and code used to conduct the analyses and generate figures in the article have been supplied as Supplemental Materials. Portions of these concepts were presented at the College on Problems of Drug Dependence 2025 conference. Demographic data are either reported elsewhere or not reported for the purposes of this secondary data analysis.

All authors have no known conflicts of interest to disclose. Mark J. Rzeszutek's time was supported by the National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health (NIH; Grant K99AA031309). Sean D. Regnier's time was supported by the National Institute on Drug Abuse, NIH (Grant K99DA060267). Some of the research reported was supported by Grant U54DA058256 (awarded to Mikhail N. Koffarnus) from the NIH and U.S. Food and Drug Administration Center for Tobacco Products. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the U.S. Food and Drug Administration. The funding sources did not have a role in writing this article or in the decision to submit it for publication.

Mark J. Rzeszutek played a lead role in conceptualization, data curation, formal analysis, methodology, software, writing–original draft, and writing–review and editing. Sean D. Regnier played a supporting role in writing–original draft and writing–review and editing. Brent A. Kaplan played a supporting role in data curation, software, and writing–review and editing. Haily K. Traxler played a supporting role in data curation and writing–review and editing. Jeffrey S. Stein played a supporting role in writing–original draft and writing–review and editing. Devin C. Tomlinson played a supporting role in writing–review and editing. Mikhail N. Koffarnus played a supporting role in writing–original draft and writing–review and editing.

Correspondence concerning this article should be addressed to Mark J. Rzeszutek or Mikhail N. Koffarnus, Department of Family and Community Medicine, College of Medicine, University of Kentucky, 2195 Harrodsburg Road, Suite 125, Lexington, KY 40504, United States. Email: mark.rzeszutek@uky.edu or koffarnus@uky.edu

Behavioral economic demand (hereafter demand), the study of commodity consumption under different constraints, has been fruitful in multiple areas of study, such as consumer decision making and substance use (Hursh & Roma, 2013, 2016; Strickland et al., 2020; Zvorsky et al., 2019). Most commonly, changes in purchasing/consumption of some commodity are assessed as a function of increase in price or some other constraint for that commodity. This is known as *own-price demand* (Hursh, 1984; Hursh & Roma, 2016). For humans, a common way to assess this is with *hypothetical purchase tasks*, which typically consist of a series of hypothetical questions that ask how much of a commodity someone would purchase and consume at various costs per unit of that commodity (Roma et al., 2016, 2017). Generally, it is expected that consumption or purchasing of a commodity will decrease as price increases (Hursh, 1984; Hursh & Silberberg, 2008). However, participant factors, like inattentiveness or misunderstanding of task instructions, may also lead to patterns that are not generally expected. Features of task design (e.g., random price order or ascending price order) may also affect response patterns. These considerations led to the development of "systematicity" criteria (Stein et al., 2015) for determining if the responses from demand tasks follow general trends expected of demand data.

The three systematicity criteria proposed by Stein et al. (2015) for demand data are (a) trend, (b) bounce, and (c) reversals from zero. The trend criterion was developed based on the assumption that consumption is typically assumed to decrease from the least constraint/cost to the highest constraint/cost. The bounce criterion was developed based on the assumption that after consumption for a commodity has decreased, consumption *ought not* increase as the constraint increases further. Finally, the reversals from zero criterion is the assumption that once a consumption of a commodity reaches zero, consumption *ought not* resume as the constraint/cost increases above the point constraint/cost suppressed all consumption. We want to stress that in cases where some criterion is violated, such as consumption not decreasing by a sufficient level as determined by an algorithm, it does not necessarily mean that data were the result of lack of participant attentiveness or actual violation of a presupposed law of demand. For example, lack of trend may be indicative of a constraint that does not affect an individual's consumption or using a range of constraints too limited to affect consumption. For example, if prices in a purchase task range between \$0 and \$10 per drink, there might be a larger proportion of trend violations compared to a range of prices between \$0 and \$100, but this would be due to the task rather than any violations of "laws of demand." While Stein et al.'s criteria account for decreases in a proportional manner based on the range of prices, the range of prices may not elicit the behavior of interest. That is, nonsystematic data may be an artifact of the methods. These originally proposed criteria, much like Johnson and Bickel's (2008) criteria inspired for delay discounting data, are not intended to be the sole justification for removing data from a data set that do not conform to hypotheses or make analyses difficult. Instead, these criteria are meant to be a tool for researchers to quantify the degree to which their methods result in unexpected data patterns and which participants were not under experimental control of the task and task instructions. Important to note, this is the same philosophy we adopt in creating the cross-commodity criteria discussed later in this article.

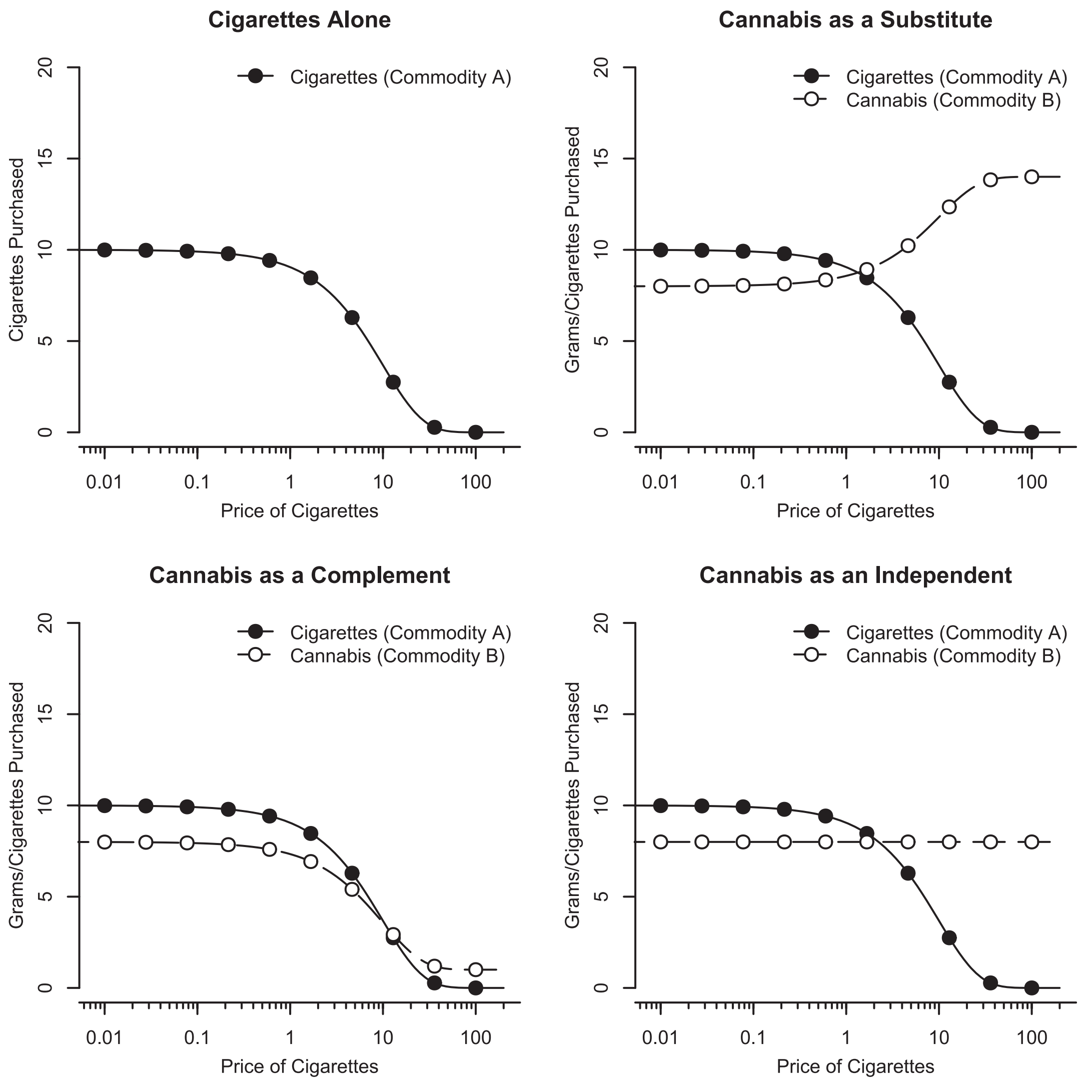
Cross-Price Demand

Now that own-price demand and own-price systematicity have been described, we introduce the concept of *cross-price demand*. Cross-price demand is the change in consumption of some commodity as a function of changes to constraints/cost of some *other* commodity. The contextualized reinforcer pathology approach partially characterizes addiction as the preference for drug rewards relative to drug and nondrug alternatives (Acuff et al., 2023), and cross-price demand can provide an important framework to understand substance misuse. That is, as the cost of Commodity A increases, cross-price demand assesses any change in consumption for a Commodity B that does not change in price. For example, an increase in the cost of cigarettes and a corresponding decrease in cigarette consumption may result in an increase, decrease, or no change in cannabis consumption (while cannabis cost remains the same). Figure 1 has examples of different patterns of changes in the consumption of cannabis (Commodity B) to increases in the cost of cigarettes (Commodity A). If cannabis consumption increases as a function of increased cigarette cost, cannabis would be considered a *substitute* for cigarettes (Figure 1: Cannabis as a Substitute). In other terms, the individual is replacing decreased cigarette consumption by switching to cannabis consumption. This may be relevant if one were interested in identifying if consumption of one commodity would replace another if it were placed under high constraint (e.g., criminalized, taxed). Conversely, if consumption of cannabis decreases as a function of increased cigarette price and a corresponding decrease in cigarette consumption, cannabis would be considered a *complement* to cigarettes (Figure 1: Cannabis as a Complement). A complementary commodity is one where its consumption decreases alongside decreases in consumption of some other commodity. Complementary relationships are useful for determining how intervening on one substance may have a beneficial decrease in another without additional intervention. Finally, cannabis would be considered an *independent* in the case that cannabis consumption does not change as a function of changes to cigarette constraints (Figure 1: Cannabis as an Independent). It is important to state that a commodity being identified as a substitute, complement, or independent is not inherent in the commodity, but instead based on the relationship between the individual/organism and their environment. Not all individuals will have the same directional changes to Commodity B, and for this, it would be relevant to consider questions like "for whom does cannabis act as a substitute for cigarettes?" rather than an absolute of "is cannabis a substitute for cigarettes?"

There has been an increased interest in examining drug consumption in the presence of alternatives (Weinsztok et al., 2023). However, as highlighted by Stein et al. (2015) and Bono et al. (2024), there are no criteria that currently exist to address if cross-commodity consumption data are "systematic." A recent review of cross-commodity purchase tasks (Weinsztok et al., 2023) found that of 23 substance-related cross-commodity purchase tasks that were published after Stein et al., about half reported criteria for data exclusion in analysis of own-price demand data (i.e., the price-manipulated commodity), and most used Stein et al.'s criteria when excluding data. However, to our knowledge, no similar criteria were applied to cross-commodity consumption data other than consumption values that were considered statistical outliers. Therefore, an algorithm to evaluate demand systematicity in cross-commodity arrangements is needed. For consistency within the article, Commodity A will always

Figure 1

Demonstrations of Own-Price and Cross-Price Demand



Note. The figure shows the idealized demand curves for own-price cigarette demand (Commodity A; top left panel) and cannabis cross-price demand (Commodity B; top right, bottom left, and bottom right panels). The *y*-axes are the number of cigarettes or grams of cannabis purchased. The *x*-axes are the cost of cigarettes. Solid line/filled circles represent the number of cigarettes purchased. Dashed line/empty circles represent the number of grams of cannabis purchased. Idealized own-price demand is modeled with Rzeszutek et al.'s (2025) equation: $Q_{\text{Own}} = Q_0 \cdot e^{(-\alpha \cdot Q_0 \cdot C)}$. Idealized cross-price demand is based on a nonlogarithmic version of the cross-price demand model (Hursh & Roma, 2013, 2016): $Q_{\text{Cross}} = Q_{\text{alone}} + I \cdot e^{(-\beta \cdot C)}$. See Rzeszutek et al. (2025) and Hursh and Roma (2016) for more details on these respective models and their parameters.

refer to the own-price demand/adjusting-price commodity whereas Commodity B will always refer to the cross-price demand/fixed-price commodity.

Determining Cross-Commodity Systematicity

Here, we offer a logical extension of the algorithm described in Stein et al. (2015) for own-price demand systematicity that is appropriate for cross-commodity consumption. We apply this to some published and unpublished cross-commodity data to demonstrate how it might be implemented. To decrease the barrier to entry for using this algorithm, we have also incorporated these analysis tools into the R package *beezdemand* (Kaplan, Gilroy, et al., 2019) and web-based tool *shinybeez* (Kaplan & Reed, 2025) as well as provide the code and data used in these analyses. Determining cross-commodity systematicity with this algorithm is a multistep process, discussed below. It is necessary to note that in the following steps and criteria, determining cross-commodity systematicity (i.e., Commodity B) is *not* dependent on the systematicity or responding of the own-price commodity (i.e., Commodity A). Because of this, responding for Commodity B may be identified as systematic, whereas responding for Commodity A may not be.

Step 1: Determine if Cross-Commodity Consumption Is a Complement, Substitute, or Independent

Cross-commodity data (i.e., responding for Commodity B) can logically take the shape of a complement, substitute, or independent, and therefore, it can uniquely increase or decrease as the cost of Commodity A increases. Therefore, cross-commodity classification (i.e., complement, substitute, independent) is necessary prior to the application of any systematicity criteria, as the classification of data series determines how the systematicity criteria will be applied. The first step is to determine which relationship is in place by calculating the overall directional change in consumption with the ΔQ criterion from Stein et al. (2015). This criterion essentially determines if there is a change in consumption, Q, from the lowest price/cost (i.e., Q_1 at C_1) and the highest price/cost (i.e., Q_n at C_n) normalized to the range of prices (see Equation 1) and compares that difference against some decision criteria based on consumption at the lowest price (e.g., $.025 \times Q_1$):

$$\Delta Q = \frac{\log_{10}(Q_1) - \log_{10}(Q_n)}{\log_{10}(C_n) - \log_{10}(C_1)}. \quad (1)$$

Much like in Stein et al.'s (2015) criteria, if ΔQ is a positive value greater than the selected threshold, then the cross-commodity consumption is flagged as a complement. That is, consumption at the lowest price is higher than consumption at the highest price. This would indicate in a gross sense that cross-commodity consumption decreased as a function of cost of the adjusting-price item. Likewise, we can flag the cross-commodity as a substitute if ΔQ in Equation 1 is a negative value that exceeds the negative threshold (e.g., −0.025). That is, cross-commodity consumption increased as a function of increasing price of the adjusting-price commodity. If cross-commodity consumption does not exceed the positive or negative threshold to be determined as a complement or substitute, respectively, then the data series is identified as an independent. To prevent problems with consumption values of zero, a small constant of 0.01 is added to each consumption value, mirroring how Stein et al.'s algorithm is implemented in *beezdemand*. Note that there is no analog of the *trend* criterion from Stein et al.'s algorithm because cross-commodity data can logically trend in any direction, contradicting the original criterion assumption that consumption typically *decreases* from the lowest to highest priced commodity.

Step 2: Determine Systematicity Based on Cross-Commodity Category

Once the cross-commodity data series (i.e., Commodity B) has been identified as either a complement, substitute, or independent based on gross trend, it can be assessed for systematicity criteria specific to its classification. The two categories of systematicity criteria we propose for cross-commodity demand are the bounce criteria and the reversals from zero/returns to zero criteria.

Bounce

After Commodity B (i.e., cross-commodity) consumption has been flagged as either a complement, substitute, or independent, a version of the bounce criterion from Stein et al. (2015) is applied to the data series. If the cross-commodity consumption was flagged as a complement, the bounce criterion from Stein et al. is identical to that of own-price consumption. For complements, this is done by determining if a response increases by 25% of Q_1 from the previous point. Note that this calculation is done using untransformed values. Then, to determine if a data series is "systematic" or "price-consistent responding," the number of total number of bounces identified in this way is compared to the number of bounces over the number of possible price increases (i.e., moving from left to right), yielding a percentage. The top left panel in Figure 2 (Own Price/Complement Systematicity) has an example of how the decision boundary is determined and which points are determined as violating the bounce criterion. For Stein et al.'s criteria, a .10 proportion of opportunities exceeded is recommended, but this can be increased or decreased based on researcher need. For example, if there are 11 prices used in a purchase task, this means there are 10 opportunities for a bounce to occur. If there are 10 opportunities, then there need to be at least two points that are determined as bouncy for the data series to be flagged for bounce using the .10 proportion threshold. Using this value also means that when using fewer than 10 price points, this .10 threshold will result in any bounces flagging a data series.

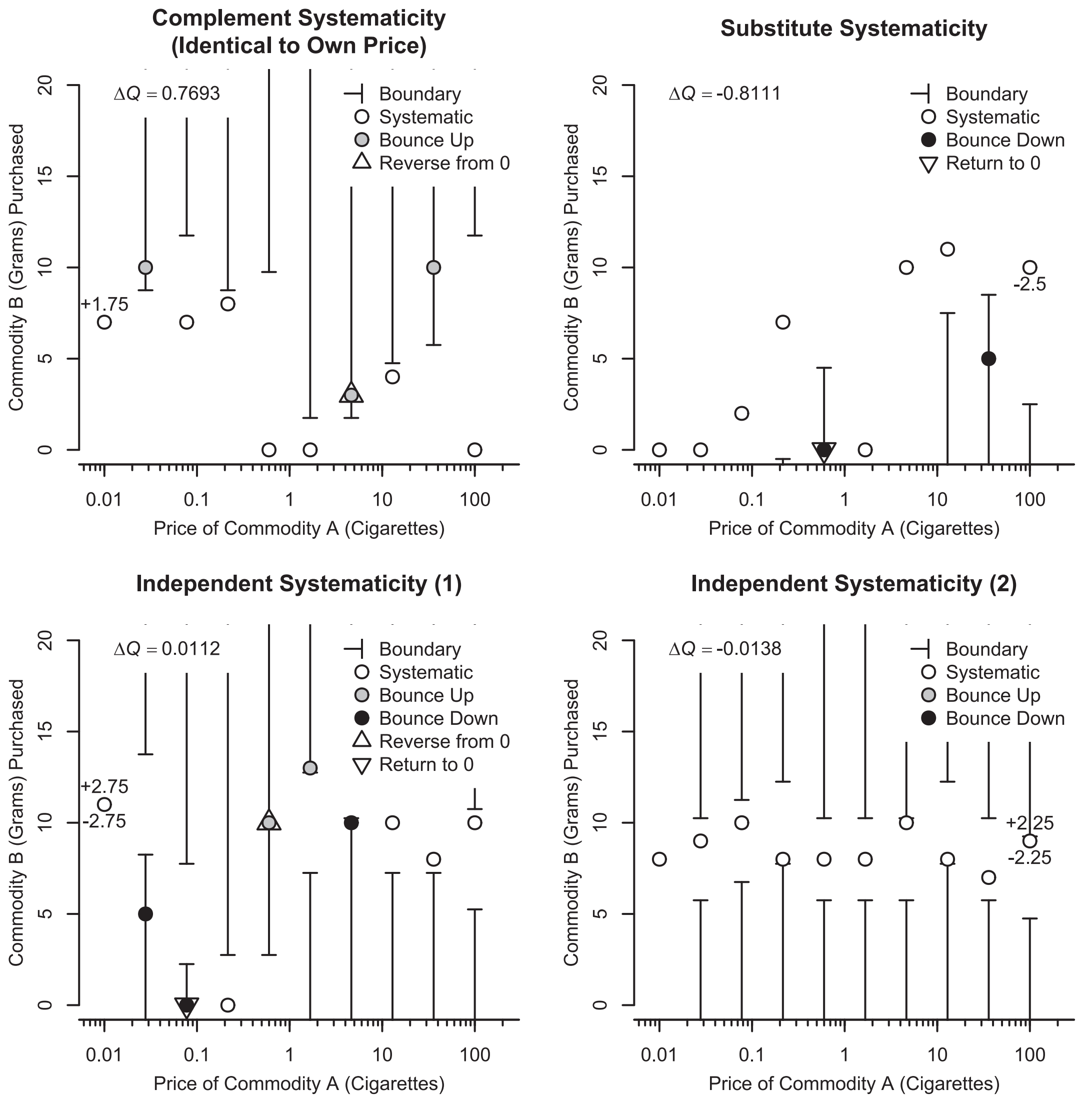
For substitutes, this is the same process but is based on a decrease in responding as price increases, using a value of 25% of Q_n (consumption at the highest price) for each subsequent purchase as the cost of the adjusting-price item increases. Consumption at the highest price is used as a basis for determining bounce size as this is generally the point of highest consumption for substitutes. The top right panel in Figure 2 (Substitute Systematicity) has an example of how this algorithm is applied for data series identified as a substitute.

For independents, the 25% threshold is calculated as 25% of the maximum of consumption at Q_0 and Q_n. Using this value, both bounces up and bounces down are calculated using the logic for complements and substitutes, respectively. To determine if a data series flagged as an independent is not "systematic," the *minimum* value of bounce violations using the complement and substitute procedure is set as the final bounce value. This is because an independent may trend toward a complement or substitute, and the criteria should not be biased against such data.[1] By taking the minimum of the complement and substitute criteria, only bounces that cannot be accounted for by a slight trend in data should be flagged. For example, in the case of the bottom left panel of Figure 2, Independent Systematicity (1), the data series flagged as independent has three instances of bounce down and two instances of bounce up. The two instances of bounce up would be used as the final count of bounce violations to determine if this data set met the 10% systematicity threshold. There is also an example of an independent data series with

[1] While the proposed minimum may be more conservative than other options, such as the mean of up and down bounce violations or using maximum bounce, alternative calculations such as these are important to explore. We have conducted an additional analysis of these different methods of identifying bounce for independents in the Supplemental Material.

Figure 2

Demonstrations of the Cross-Commodity Systematicity Algorithm



Note. In all panels, the *y*-axes are the grams of cannabis purchased (Commodity B) and the *x*-axes are the cost of cigarettes (Commodity A). Points that are "systematic" as determined by the algorithm are white circles, and "nonsystematic" points are indicated with gray fill (bounce up), black fill (bounce down), triangles up (reversal from zero), or triangles down (return to zero). Points can be more than one nonsystematic category. Decision boundaries are indicated by solid vertical lines with flat tops/bottoms. Data points with a value above, below, or both are the points used to determine the range of the boundary and, dependent on if the data path was determined as complement, substitute, or independent, the direction of the boundary. While either the first or last point of the data series can be used to determine the boundary, all bounce violations are determined from left to right.

variability not meeting the bounce threshold in either direction, which can be found in the bottom right panel of Figure 2, Independent Systematicity (2). In all cases, complements, substitutes, or independents, *the bounce criterion is always applied from left to right*. The rationale for this is that the definition for all own-price and cross-price consumption is based on the following assumption: *As the cost of the own-price commodity increases, the consumption of the cross-price commodity is affected*.

Reversals From Zero and Returns to Zero

We have also incorporated the reversals from zero criteria much like Stein et al. (2015), where a data point is flagged as having "reversing from zero" when a response follows a given number of zeros. For data identified as a complement, the implementation of this criterion is identical to Stein et al.'s reversals from zero criteria, where the default is a reversal being flagged after a value greater than zero is present after two consecutive zero responses. To make a similar consideration for substitutes, we introduce the criteria "returns to zero," where if a data series has been identified as a substitute (i.e., trends up as price increases), then any responses of zero following two nonzero responses is flagged as a return to zero. In the case of independents, both reversals from and returns to zero are calculated and the *minimum* value of the two is used to determine systematicity. Examples of reversals from and returns to zero can be found in Figure 2.

Step 3: Report Own-Price and Cross-Price Systematicity

When applying behavioral economic demand systematicity criteria to a data series, rates of systematicity should be reported in manuscripts. This allows for meta-analyses to determine if there are systematic deviations from systematicity in different groups, tasks, or commodities. This allows for transparency within behavioral economic demand data and can help to improve task development. See Tables 1–3 as examples.

Step 4: Sensitivity Analyses and Data Exclusions

This step is extremely important for own-price and cross-price demand. A variety of different criteria have been used to identify nonsystematic data series (e.g., Bruner & Johnson, 2014; Kaplan et al., 2018; Stein et al., 2015; Weinsztok et al., 2023) and other illogical data patterns (e.g., extreme consumption values), which may then result in either removal of data series or replacement of specific consumption values (e.g., Winsorizing). *While the systematicity criteria presented here and in* Stein et al. (2015) *are intended to be descriptive in nature*, if any data are excluded from an analysis based on any criteria, we recommend that *at a minimum* the criteria for exclusion be clearly articulated and described and that a sensitivity analysis be conducted for the data with exclusions and the data without exclusions. To conduct a sensitivity analysis in this case, one would run the same analysis on a data set before and after data were excluded to determine whether the removal of data affected the conclusions of the study. Whether the conclusions were affected should then be reported in the article. The reason why this is such an important step is that depending on the nature of the exclusions, results could be adversely affected and conclusions could be fundamentally different. However, if these sensitivity analyses are not conducted, it is impossible to know how interpretations may be affected. If a sensitivity analysis cannot be conducted, reasons should be explicit as to why. For example, if a proposed nonlinear analysis results in nonconvergence issues due to the irregularity in the data, this should also be clearly stated. Also, if possible, alternate endpoints other than demand used in the study or participant characteristics (e.g., use disorder severity, age, sex) should be compared before and after exclusion to detect whether the exclusion biased the sample in other ways.

Example Applications of Cross-Commodity Systematicity Criteria

To demonstrate these proposed cross-commodity systematicity criteria, we present some cross-commodity demand data from three different data sets. The first data set is from already published data (Kaplan, Pope, et al., 2019). These consist of between-subject cross-commodity consumption for cigarettes and alternative cigarettes over three different narrative scenarios from participants recruited from Amazon Mechanical Turk (MTurk) in 2017. This data set had 16 price points for the adjusting-price cigarettes. The second data set is unpublished data of within-subject cross-commodity demand between cigarettes and cannabis consisting of participants recruited from MTurk in 2023. This data set had 11 prices for adjusting-price cannabis and 14 prices for adjusting-price cigarettes. The third data set is from an ongoing study that incorporates the Experimental Tobacco Marketplace (ETM) consisting of participants recruited from the Appalachian region in Kentucky (Clinical Trials ID: NCT06234722). This data set has six prices for adjusting-price

Table 1

Low Nicotine—Kaplan et al.'s (2018) Systematicity Data

	Own-price/Commodity A systematicity violation		
Commodity (n)	Trend/ΔQ	B-Up	Reversals
Own 100% (67)	1 (1.5%)	3 (4.5%)	2 (3.0%)
Own 2% (57)	3 (5.3%)	0 (0%)	3 (5.3%)
Own 98% (59)	1 (1.7%)	1 (1.7%)	0 (0%)

	Cross-price/Commodity B systematicity metric								
	Cross-commodity category			Systematicity violation					
Commodity (n)	ΔQ_{Down}	ΔQ_{Up}	ΔQ_{None}	B-Up[a]	B-Down[a]	B-Ind[a]	B-Any	Reversals	Returns
Cross 100% (67)	9 (13.4%)	24 (35.8%)	34 (50.7%)	0 (0%)	1 (4.2%)	1 (2.9%)	2 (3%)	6 (9%)	1 (1.5%)
Cross 2% (57)	6 (10.5%)	15 (26.3%)	36 (63.2%)	1 (16.7%)	2 (13.3%)	2 (5.6%)	5 (8.8%)	3 (5.3%)	0 (0%)
Cross 98% (59)	10 (16.9%)	26 (44.1%)	23 (39%)	0 (0%)	2 (7.7%)	3 (13%)	5 (8.5%)	4 (6.8%)	1 (1.7%)

Note. Own-price systematicity violations are based on Stein et al.'s (2015) criteria and cross-price systematicity on the proposed criteria. Count and percentages by commodity across own-price/adjusting commodity and cross-price/fixed commodity by condition. Trend/ΔQ = Stein et al.'s trend criteria violations; ΔQ_{Down} = trend identified down (complement); ΔQ_{Up} = trend identified up (substitute); ΔQ_{None} = trend identified as neither up nor down (independent); B-Up = bounce up violations; B-Down = bounce down violations; B-Ind = bounces up and down violations for independents; B-Any = overall bounces irrespective of trend; Reversals = reversals from zero; Returns = returns to zero.

[a] For cross-commodity bounce violations, percentages are based on the number of trends identified for the relevant bounce criteria.

Table 2
Unpublished Cannabis and Cigarette Systematicity Data

Commodity (*n*)	Own-price/Commodity A systematicity violation: Trend/ΔQ	B-Up	Reversals
Own Can. (99)	42 (42.4%)	40 (40.4%)	1 (1%)
Own Cig. (99)	37 (37.4%)	34 (34.3%)	0 (0%)

	Cross-price/Commodity B systematicity metric								
	Cross-commodity category			Systematicity violation					
Commodity (*n*)	ΔQ_{Down}	ΔQ_{Up}	ΔQ_{None}	B-Up[a]	B-Down[a]	B-Ind[a]	B-Any	Reversals	Returns
Cross Cig. (99)	40 (40.4%)	20 (20.2%)	39 (39.4%)	7 (17.5%)	6 (30%)	9 (23.1%)	22 (22.2%)	1 (1%)	0 (0%)
Cross Can. (99)	39 (39.4%)	19 (19.2%)	41 (41.4%)	13 (33.3%)	11 (57.9%)	8 (19.5%)	32 (32.3%)	0 (0%)	1 (1%)

Note. Own-price systematicity violations are based on Stein et al.'s (2015) criteria and cross-price systematicity on the proposed criteria. Count and percentages by commodity across own-price/adjusting commodity and cross-price/fixed commodity by condition. Own Can. = own-price cannabis demand; Own Cig. = own-price cigarette demand; Cross Cig. = cross-price cigarette demand; Cross Can. = cross-price cannabis demand; Trend/ΔQ = Stein et al.'s trend criteria violations; ΔQ_{Down} = trend identified down (complement); ΔQ_{Up} = trend identified up (substitute); ΔQ_{None} = trend identified as neither up nor down (independent); B-Up = bounce up violations; B-Down = bounce down violations; B-Ind = bounces up and down violations for independents; B-Any = overall bounces irrespective of trend; Reversals = reversals from zero; Returns = returns to zero.
[a] For cross-commodity bounce violations, percentages are based on the number of trends identified for the relevant bounce criteria.

preferred-flavor cigarettes. While more products were assessed in the ETM, we only looked at cigarette smokers and their own-price demand for preferred-flavor cigarettes, cross-commodity purchasing of nonpreferred cigarettes, and cross-commodity purchasing of e-cigarettes. In all three data sets being used for demonstrative purposes, prices were presented in ascending order. For the low-nicotine study, all purchasing opportunities were presented on the same page, whereas for the cannabis/cigarette and ongoing ETM data, purchasing opportunities were presented on separate pages.

Data Analysis

For the own-price demand data series, Stein et al.'s (2015) criteria for trend (i.e., Trend/ΔQ), bounces up, and reversals from zeros were applied. For cross-price data series, Steps 1–3 described above were applied. No additional screening that occurred in Kaplan, Pope, et al. (2019) was used to exclude data prior to this analysis. Code and data used for the analysis are available in the Supplemental Materials. Analyses were conducted in R 4.4.2 (R Core Team, 2024), and the packages *data.table* (Dowle & Srinivasan, 2020) and *beezdemand* (Kaplan, Gilroy, et al., 2019) were used for data organization and basing cross-commodity systematicity, respectively.

We also conducted an analysis of different bounce thresholds for cross-commodity categorization across different ΔQ rates. The purpose of this was to demonstrate how changing these decision thresholds affects the relative proportion of data paths that were identified and how nonsystematic data changed as a function of this. The ΔQ category thresholds assessed were .001, .025, .075, .150, .250, .375, and .500. The bounce thresholds assessed were .050, .100, .200, .300, and .400. We also conducted the same analysis with own-price demand from these data sets, which is further described and reported in the Supplemental Material.

Table 3
In Progress Experimental Tobacco Marketplace Data

Commodity (*n*)	Own-price/Commodity A systematicity violation: Trend/ΔQ	B-Up	Reversals
Own Cig. (47)	3 (6.4%)	5 (10.6%)	0 (0%)

	Cross-price/Commodity B systematicity metric								
	Cross-commodity category			Systematicity violation					
Commodity (*n*)	ΔQ_{Down}	ΔQ_{Up}	ΔQ_{None}	B-Up[a]	B-Down[a]	B-Ind[a]	B-Any	Reversals	Returns
Cross Cig. (47)	4 (8.5%)	6 (12.8%)	37 (78.7%)	3 (75%)	0 (0%)	2 (5.4%)	5 (10.6%)	1 (2.1%)	1 (2.1%)
Cross E-Cig. (47)	6 (12.8%)	15 (31.9%)	26 (55.3%)	3 (50%)	3 (20%)	2 (7.7%)	8 (17%)	2 (4.3%)	1 (2.1%)

Note. Own-price systematicity violations are based on Stein et al.'s (2015) criteria and cross-price systematicity on the proposed criteria. Count and percentages by commodity across own-price/adjusting commodity and cross-price/fixed commodity by condition. Own Cig. = own-price cigarette demand; Cross Cig. = cross-price cigarette demand; Cross E-Cig. = cross-price e-cigarette demand; Trend/ΔQ = Stein et al.'s trend criteria violations; ΔQ_{Down} = trend identified down (complement); ΔQ_{Up} = trend identified up (substitute); ΔQ_{None} = trend identified as neither up nor down (independent); B-Up = bounce up violations; B-Down = bounce down violations; B-Ind = bounces up and down violations for independents; B-Any = overall bounces irrespective of trend; Reversals = reversals from zero; Returns = returns to zero.
[a] For cross-commodity bounce violations, percentages are based on the number of trends identified for the relevant bounce criteria.

Results

Tables 1–3 contain the counts and percentages of own-price and cross-price demand for Kaplan, Pope, et al.'s (2019) data set, the cigarette and cannabis data set, and the ETM data set, respectively. For bounce criteria, because the number of price points differed, this also results in the threshold that was used (.10 proportion of opportunities) to be more sensitive. That is, for the ETM data set, there are only five opportunities for a bounce to occur, and thus a single bounce would be above the threshold. In contrast, for Kaplan, Pope, et al.'s study, there were 15 opportunities for bounce to occur, which required two bounces to meet/exceed the .10 proportion for being identified as having failed the bounce criteria. The cigarette and cannabis data set also required at least two points being identified as a bounce to be flagged. For own-price demand, most participants from Kaplan, Pope, et al.'s and the ongoing ETM study had data series that would pass the trend criteria (>90% passed) and similar numbers for the bounce and reversals criteria. By contrast, the cigarette and cannabis own-price demand only had ∼60% of participants who displayed a decreasing trend for either commodity. Bounce for both cigarette and cannabis from this data set also was substantially higher than the other two, with ∼40% of participants' data series being flagged for having exceeded the bounce threshold.

Cross-commodity trends were, from lowest to highest, complements, substitutes, and independents as determined by the ΔQ assignment (see the "default" value proportions in Figure 3). Generally, the highest number of bounce violations occurred for complements, while the substitutes and independents had lower number of bounce violations. For the cigarette and cannabis data set, most data series were identified as complements or independents ∼40% for both fixed-price cigarettes and cannabis, with roughly 20% of participants' data series being identified as substitutes. Bounce violations were also generally higher in the cigarette and cannabis data set for cross-commodity demand as well. Results of the overall threshold analyses are in Figure 3, and results from individual data sets used in these analyses are in the Supplemental Material. Cross-commodity categorization and bounce violations using "default" values of ΔQ (0.025) and bounce (.10) are highlighted with thick black squares in this figure. As the ΔQ threshold for cross-commodity assignment increased, the relative proportion of data series identified as independent increased while substitutes and complements decreased. Results of the threshold analysis for own-price and cross-price systematicity for each study can be found in the Supplemental Material.

Discussion

The purpose of this study was to propose and demonstrate the use of cross-commodity systematicity criteria, expanding upon Stein et al.'s (2015) own-price demand criteria. When possible, we attempted to create rules that are consistent with Stein et al.'s criteria, only adding rules and stipulations when necessary to accommodate the different patterns that cross-commodity demand data can coherently take. As cross-commodity data can reasonably show an increasing trend, a decreasing trend, or no trend, there is no "trend" criterion in this version. The "bounce" and "reversals from zero" criteria were then adapted to accommodate different patterns of cross-commodity data.

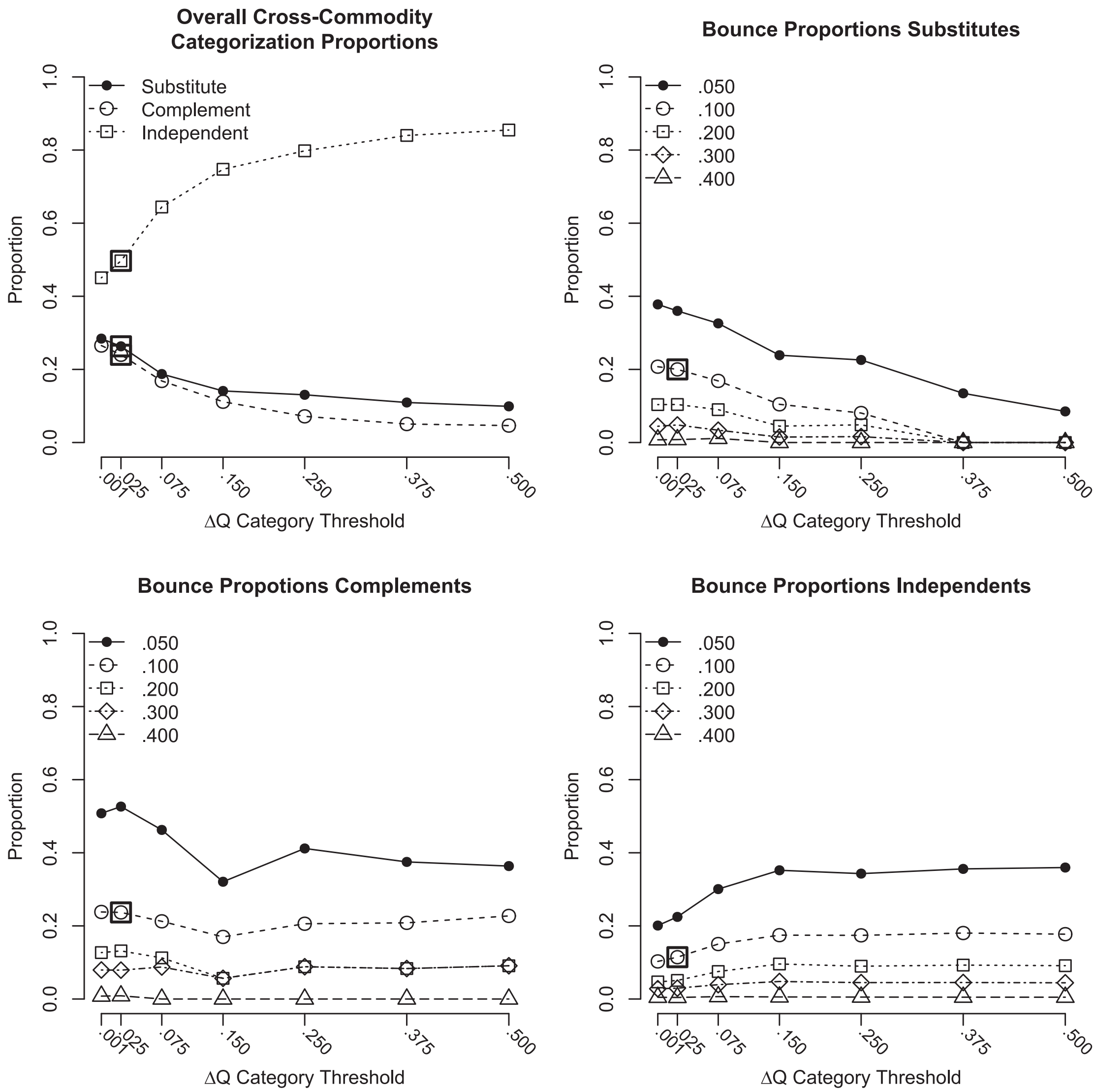
The goal of these criteria is to provide a standardized way to assess the systematicity of cross-commodity demand data. These criteria could be useful to characterize data sets, revealing which participants were under experimental control of the task. Additionally, these criteria could be used as a dependent measure when developing new demand tasks or refining experimental procedures to assess which set of procedures results in the most systematic data. For example, one could investigate whether the implementation of an instructional demand task session with a researcher-led question-and-answer period improves the systematicity of demand data compared to standard written instructions (for examples of instructions on other data quality metrics, see Freitas-Lemos et al., 2022).

An additional use of these criteria that should be pursued with caution is to screen data for elimination from further data analysis. Common analyses for demand data typically involve either (a) fitting all data to a single mixed-effects model or (b) extracting demand parameters from individual subjects' demand curves and then analyzing further (i.e., a two-stage analysis). The presence of nonsystematic data can pose problems for either method. In mixed-effects models, nonsystematic data can result in the model failing to converge, yielding no result. In a two-stage analysis, it can be impossible to estimate demand parameters from nonsystematic data that do not conform to the assumptions of demand. However, removal of data that do not conform with expected results can bias the results toward the researcher's hypothesis and violate the assumptions of statistical testing. If data are removed from a data set in this way, a prudent step is to perform a sensitivity analysis to determine if the removal affected the conclusions of the study, followed by transparent reporting of the sensitivity analysis in any article or presentation. When possible though, a more statistically sound method could be to report the prevalence of systematicity, but then include all data in the final data analysis with a method that allows for the incorporation of nonsystematic data. Mixed-effects models can allow for the incorporation of all demand data in a single analysis, even if some data do not conform to the expected data patterns or are nonsystematic. Such methods preserve the assumptions of statistical tests that rely on estimating the variance present in the whole data set, and do not artificially reduce variance by removing those subjects with the highest deviation from group means. For a detailed explanation of the use of mixed-effects modeling for demand data, see Kaplan et al. (2021).

The criteria described here include two systematicity criteria for cross-commodity demand data, but the presence of these criteria does not imply that other criteria may not exist and be relevant in specific cases. One additional criterion that we (e.g., Rzeszutek et al., 2022, 2023) and others (e.g., Dwyer et al., 2022) have used in the past as an indicator of participants not being under experimental control of the task is the presence of extreme consumption values that are not feasible given the task constraints. For example, a common set of cigarette purchase task instructions asks participants to indicate how many cigarettes they would purchase and consume in a 24-hr period at various prices (Reed et al., 2020). If a participant follows these instructions, it would be likely be physically impossible and therefore extremely unlikely to consume more than ∼10 packs of cigarettes (200 cigarettes) in one 24-hr period. However, some participants who are not under experimental control of these instructions may indicate that they would consume more than 200 cigarettes, sometimes fantastical amounts exceeding 1,000 cigarettes. The presence of extreme values in a data set such as this can have a large impact on statistical results and the estimation of group means. Therefore, like in other areas, some researchers have used this as a basis for removal of data or "Winsorizing" data to values no greater than 3.29 standard deviations from the mean.

Figure 3

Cross-Commodity Threshold Analyses



Note. The figure shows the proportions of cross-commodity identifications/categorizations based on different ΔQ thresholds and bounce thresholds for all three data sets; *y*-axes, proportions, *x*-axes, ΔQ thresholds. For the top left panel, data paths represent the proportions of cross-commodity category for all three data sets. For all other panels, data paths represent the proportions of data series that were identified to have a bounce based on different bounce thresholds. Thick squares, which highlight data points, represent the "default" criteria of a ΔQ of .025 and a bounce threshold of .1.

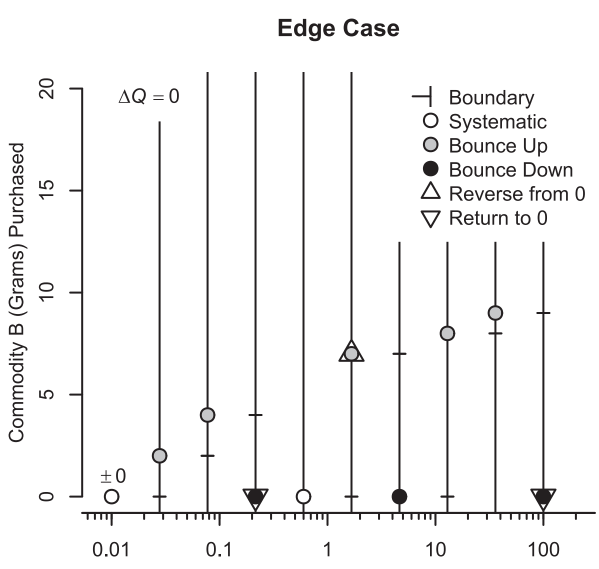
While the criteria and analyses conducted for this article cannot be used to identify "best practice" for conducting cross-commodity studies to decrease nonsystematic data, we can assume that certain presentations may decrease nonsystematic responding. For example, randomized sequences appear to increase nonsystematic responding with hypothetical purchase tasks (Salzer et al., 2021; Tomlinson et al., 2023). We cannot definitively assume that factors that affect own-price demand systematicity also affect cross-commodity systematicity, but it is likely that best practices from own-price demand may also "improve" cross-commodity demand. Further examination of how cross-commodity demand is affected by methodological implementations is warranted.

There exist limitations of this proposed algorithm. First, the cutoff points we used as recommendations here for defining substitutes, complements, and the number of bounces that indicate a deviation from systematicity are arbitrary. Because of this, we have conducted threshold analyses to compare how different ΔQ categorizations and bounce criteria may affect cross-commodity systematicity decisions. While the current article focuses on directly applying Stein et al.'s (2015) criteria and thresholds to cross-commodity demand, we want to emphasize that the general standards here may not be universally applicable.[2] There could be no universal cutoff that applies in all cases, and researchers may decide that different cutoff values are more appropriate for their experimental conditions. However, the presence of a standard across studies can improve reproducibility and interpretability of studies across time and environments. Second, the criteria proposed here may not fully quantify the degree of systematicity in all cases. Edge cases will exist where the algorithm results in flagging a greater or fewer number of data points than what visually appear to be nonsystematic. For example, a possible edge case exists if both the first and last data points in a series are at zero consumption. Because the algorithm relies on the greater of these two points to determine the threshold for determining a "bounce," both of these values being zero results in any nonzero change in consumption from one point to the next being classified as a "bounce" (Figure 4). Also, depending on the units of measurement being used (e.g., packs of cigarettes vs. individual cigarettes), this could also result in higher rates of bounce violations.

In conclusion, the criteria presented here may be used to quantify the degree of systematicity in cross-commodity demand data to help researchers assess experimental conditions to improve experimental control and to flag specific data series that violate the assumptions of demand. We recommend that these criteria be used judiciously in accordance with statistical norms to characterize their data and, if necessary and only after appropriate statistical measures such as sensitivity analysis have been conducted, flag data for removal or separate analyses.

Figure 4

Possible Edge Case Affecting Cross-Commodity Identification



Note. The *y*-axis is the grams of cannabis purchased (Commodity B), and the *x*-axis is the cost of cigarettes (Commodity A). Points that are "systematic" as determined by the algorithm are white circles, and "nonsystematic" points are indicated with gray fill (bounce up), black fill (bounce down), triangles up (reversal from zero), or triangles down (return to zero). Points can be more than one nonsystematic category. Decision boundaries are indicated by solid vertical lines with flat tops/bottoms. Data points with a value above, below, or both are the points used to determine the range of the boundary and, dependent on if the data path was determined as complement, substitute, or independent, the direction of the boundary.

[2] While we only assessed these criteria across three data sets, results of the threshold analysis seem to result in outcomes that are commensurate with Stein et al.'s (2015) criteria on own-price data. Assessment of these nonsystematic algorithms and thresholds across many more data sets is warranted to determine "best practices" for cross-commodity demand data.

References

Acuff, S. F., MacKillop, J., & Murphy, J. G. (2023). A contextualized reinforcer pathology approach to addiction. *Nature Reviews Psychology*, *2*(5), 309–323. https://doi.org/10.1038/s44159-023-00167-y

Bono, R. S., White, A. M., Hoetger, C., Lipato, T., Bickel, W. K., Cobb, C. O., & Barnes, A. J. (2024). Expanding on cross-price elasticity: Understanding tobacco product demand and substitution from the cross-price purchase task. *Journal of the Experimental Analysis of Behavior*, *121*(2), 175–188. https://doi.org/10.1002/jeab.890

Bruner, N. R., & Johnson, M. W. (2014). Demand curves for hypothetical cocaine in cocaine-dependent individuals. *Psychopharmacology*, *231*(5), 889–897. https://doi.org/10.1007/s00213-013-3312-5

Dowle, M., & Srinivasan, A. (2020). *data.table: Extension of "data.frame"* (Version 1.13.0) [Computer software]. https://CRAN.R-project.org/package=data.table

Dwyer, M. J., Burrows, C. A., Drossel, C., Raiff, B. R., & Kirby, K. C. (2022). Using the novel functional purchase task to examine prescription stimulant drug effect preferences in college students. *Experimental and Clinical Psychopharmacology*, *30*(6), 774–786. https://doi.org/10.1037/pha0000539

Freitas-Lemos, R., Tegge, A. N., Craft, W. H., Tomlinson, D. C., Stein, J. S., & Bickel, W. K. (2022). Understanding data quality: Instructional comprehension as a practical metric in crowdsourced investigations of behavioral economic cigarette demand. *Experimental and Clinical Psychopharmacology*, *30*(4), 415–423. https://doi.org/10.1037/pha0000579

Hursh, S. R. (1984). Behavioral economics. *Journal of the Experimental Analysis of Behavior*, *42*(3), 435–452. https://doi.org/10.1901/jeab.1984.42-435

Hursh, S. R., & Roma, P. G. (2013). Behavioral economics and empirical public policy. *Journal of the Experimental Analysis of Behavior*, *99*(1), 98–124. https://doi.org/10.1002/jeab.7

Hursh, S. R., & Roma, P. G. (2016). Behavioral economics and the analysis of consumption and choice. *Managerial and Decision Economics*, *37*(4–5), 224–238. https://doi.org/10.1002/mde.2724

Hursh, S. R., & Silberberg, A. (2008). Economic demand and essential value. *Psychological Review*, *115*(1), 186–198. https://doi.org/10.1037/0033-295X.115.1.186

Johnson, M. W., & Bickel, W. K. (2008). An algorithm for identifying nonsystematic delay-discounting data. *Experimental and Clinical Psychopharmacology*, *16*(3), 264–274. https://doi.org/10.1037/1064-1297.16.3.264

Kaplan, B. A., Foster, R. N. S., Reed, D. D., Amlung, M., Murphy, J. G., & MacKillop, J. (2018). Understanding alcohol motivation using the alcohol purchase task: A methodological systematic review. *Drug and Alcohol Dependence*, *191*, 117–140. https://doi.org/10.1016/j.drugalcdep.2018.06.029

Kaplan, B. A., Franck, C. T., McKee, K., Gilroy, S. P., & Koffarnus, M. N. (2021). Applying mixed-effects modeling to behavioral economic demand: An introduction. *Perspectives on Behavior Science*, *44*(2), 333–358. https://doi.org/10.1007/s40614-021-00299-7

Kaplan, B. A., Gilroy, S. P., Reed, D. D., Koffarnus, M. N., & Hursh, S. R. (2019). The R package *beezdemand*: Behavioral economic easy demand.

Perspectives on Behavior Science, *42*(1), 163–180. https://doi.org/10.1007/s40614-018-00187-7

Kaplan, B. A., Pope, D. A., DeHart, W. B., Stein, J. S., Bickel, W. K., & Koffarnus, M. N. (2019). Estimating uptake for reduced-nicotine cigarettes using behavioral economics. *Tobacco Regulatory Science*, *5*(3), 264–279. https://doi.org/10.18001/TRS.5.3.5

Kaplan, B. A., & Reed, D. D. (2025). shinybeez: A Shiny app for behavioral economic easy demand and discounting. *Journal of the Experimental Analysis of Behavior*, *123*(2), 355–376. https://doi.org/10.1002/jeab.70000

R Core Team. (2024). *R: A language for statistical computing* (Version 4.4.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Reed, D. D., Naudé, G. P., Salzer, A. R., Peper, M., Monroe-Gulick, A. L., Gelino, B. W., Harsin, J. D., Foster, R. N. S., Nighbor, T. D., Kaplan, B. A., Koffarnus, M. N., & Higgins, S. T. (2020). Behavioral economic measurement of cigarette demand: A descriptive review of published approaches to the cigarette purchase task. *Experimental and Clinical Psychopharmacology*, *28*(6), 688–705. https://doi.org/10.1037/pha0000347

Roma, P. G., Hursh, S. R., & Hudja, S. (2016). Hypothetical purchase task questionnaires for behavioral economic assessments of value and motivation. *Managerial and Decision Economics*, *37*(4–5), 306–323. https://doi.org/10.1002/mde.2718

Roma, P. G., Reed, D. D., DiGennaro Reed, F. D., & Hursh, S. R. (2017). Progress of and prospects for hypothetical purchase task questionnaires in consumer behavior analysis and public policy. *The Behavior Analyst*, *40*(2), 329–342. https://doi.org/10.1007/s40614-017-0100-2

Rzeszutek, M. J., Gipson-Reichardt, C. D., Kaplan, B. A., & Koffarnus, M. N. (2022). Using crowdsourcing to study the differential effects of cross-drug withdrawal for cigarettes and opioids in a behavioral economic demand framework. *Experimental and Clinical Psychopharmacology*, *30*(4), 452–465. https://doi.org/10.1037/pha0000558

Rzeszutek, M. J., Kaplan, B. A., Traxler, H. K., Franck, C. T., & Koffarnus, M. N. (2023). Hyperbolic discounting and exponentiated demand: Modeling demand for cigarettes in three dimensions. *Journal of the Experimental Analysis of Behavior*, *119*, 169–191. https://doi.org/10.1002/jeab.818

Rzeszutek, M. J., Regnier, S. D., Franck, C. T., & Koffarnus, M. N. (2025). Overviewing the exponential model of demand and introducing a simplification that solves issues of span, scale, and zeros. *Experimental and Clinical Psychopharmacology*, *33*(2), 209–223. https://doi.org/10.1037/pha0000754

Salzer, A. R., Strickland, J. C., Stoops, W. W., & Reed, D. D. (2021). An evaluation of fixed and randomized price sequence on the alcohol purchase task. *Experimental and Clinical Psychopharmacology*, *29*(4), 295–301. https://doi.org/10.1037/pha0000421

Stein, J. S., Koffarnus, M. N., Snider, S. E., Quisenberry, A. J., & Bickel, W. K. (2015). Identification and management of nonsystematic purchase task data: Toward best practice. *Experimental and Clinical Psychopharmacology*, *23*(5), 377–386. https://doi.org/10.1037/pha0000020

Strickland, J. C., Campbell, E. M., Lile, J. A., & Stoops, W. W. (2020). Utilizing the commodity purchase task to evaluate behavioral economic demand for illicit substances: A review and meta-analysis. *Addiction*, *115*(3), 393–406. https://doi.org/10.1111/add.14792

Tomlinson, D. C., Freitas-Lemos, R., Tegge, A. N., Yeh, Y.-H., Dwyer, C. L., Stein, J. S., & Bickel, W. K. (2023). Presentation matters: Effects of cigarette purchase task design on systematic data and purchasing behavior. *Experimental and Clinical Psychopharmacology*, *31*(6), 1017–1022. https://doi.org/10.1037/pha0000643

Weinsztok, S. C., Reed, D. D., & Amlung, M. (2023). Substance-related cross-commodity purchase tasks: A systematic review. *Psychology of Addictive Behaviors*, *37*(1), 72–86. https://doi.org/10.1037/adb0000851

Zvorsky, I., Nighbor, T. D., Kurti, A. N., DeSarno, M., Naudé, G., Reed, D. D., & Higgins, S. T. (2019). Sensitivity of hypothetical purchase task indices when studying substance use: A systematic literature review. *Preventive Medicine*, *128*, Article 105789. https://doi.org/10.1016/j.ypmed.2019.105789

Received September 19, 2025
Revision received December 14, 2025
Accepted January 9, 2026 ■