Adaptive Purchase Tasks in the Operant Demand Framework

Shawn P. Gilroy

Louisiana State University

Mark J. Rzeszutek

University of Kentucky

Mikhail N. Koffarnus

University of Kentucky

Derek D. Reed

Institutes for Behavior Resources, Inc

Steven R. Hursh

Institutes for Behavior Resources, Inc

Department of Psychiatry and Behavioral Science,

Johns Hopkins University School of Medicine

Author Note:

We have no known conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Shawn P. Gilroy, 220 Audubon Hall, Louisiana State University, Baton Rouge, Louisiana 70803. Email: sgilroy1@lsu.edu

Abstract

Various avenues exist for quantifying the effects of reinforcers on behavior. Numerous nonlinear models derived from the framework of Hursh and Silberburg (2008) are often applied to elucidate key metrics in the operant demand framework (e.g., Q_0, P_{MAX}), each approach having respective strengths and tradeoffs. This work presents and evaluates a model-free approach to elucidating key features of reinforcer demand using an adaptive task as opposed to deriving them from modeling applied to data using fixed price assays and calculus. An adaptive algorithm for hypothetical purchase tasks based on Reinforcement Learning is presented and evaluated for use in elucidating individual-level estimates of peak work (e.g., P_{MAX}). The algorithm was evaluated across 4 different iteration lengths (i.e., 5, 10, 15, and 20 questions) and equivalence tests with simulated agent responses revealed that tasks with ten or more sequentially updated questions recovered P_{MAX} values that were statistically equivalent to seeded values. Preliminary findings regarding adaptive purchasing tasks suggest that quantitative modeling may not be necessary in all applications of the operant demand framework and existing empirical metrics may be much more representative when tasks adapt in a process of exploring maximal personal utility. A short discussion on future extensions and conditions under which nonlinear modeling may or may not be necessary is presented.

Keywords: *operant demand, purchase tasks, behavioral economics*

Adaptive Purchase Tasks in the Operant Demand Framework

The Operant Demand Framework is a mature and active collection of methods designed to evaluate how the effects of reinforcers on behavior scale as a function of various ecological factors, such as 'cost' and the availability of potential alternatives (Hursh, 1980, 1984). Hursh & Silberberg (2008) emphasized the need to expand upon earlier efforts to quantify reinforcer effects (e.g., breakpoints, response rates) and move toward another reinforcer-based framework that was sensitive to the continuous nature of reinforcer effects (e.g., varying effect across potential schedules) as well as the various forms of reinforcer-reinforcer relations (e.g., complements, substitutes) that likely influence choice behavior (Hursh & Bauman, 1987; Madden et al., 2007).

The most well-represented approach to quantifying reinforcer consumption using economic principles has been the Exponential Model of Operant Demand proposed by Hursh & Silberberg (2008). This model (Eq. 1) characterizes the effects of price (P) on levels of reinforcer consumption (Q). A total of three parameters are featured in the model and inferences drawn from this model emphasize two key measures related to consumption: Demand intensity (Q_0) and the scaling of reinforcer effects as a function of Price (i.e., α and P_{MAX}).[1]

$$\log_{10} Q_0 + k(e^{\alpha Q_0 P} - 1) \quad (1)$$

Demand intensity reflects consumption when scaling due to P is zeroed out (i.e., $Q_0 = Q$ at a P of 0) and therefore reflects a dimension of reinforcer consumption independent of price. Alternatively, consumption scaled as a function of price can be both directly and indirectly estimated in the model via α and P_{MAX}, respectively. The α parameter reflects the rate of change

[1] Note. Quantity P_{MAX} is not fitted directly and is instead derived from overall model predictions; however, controlling for all other parameters, parameter α is most representative of price scaling effects.

in price elasticity (η) observed across prices, given units inferred by the intercept (Q_0) and span of the demand curve (k), and this reveals the distance in terms of P wherein the η of demand shifts from inelastic demand to elastic demand (i.e., P_{MAX}). That is, α may be viewed as the rate by which the inelastic demand for reinforcers observed at lower prices advances toward elastic demand at increased prices. Unit elasticity (i.e., η = -1) and price associated with peak work, P_{MAX}, equates to parameter α when accounting for parameters Q_0 and k (Gilroy et al., 2019). Related to P_{MAX}, the quantity O_{MAX} represents refers to the total overall level of consumption observed at P_{MAX}. These metrics have been found to each capture distinct and meaningful dimensions of reinforcer consumption for various types of reinforcer consumption (Aston et al., 2017; Bidwell et al., 2012; Mackillop et al., 2009).

Critical Elements in the Operant Demand Framework

Individual patterns of consumption have been linked to various ecological and contextual factors (Strickland et al., 2022). Researchers applying research synthesis to characterize various forms of substance use (e.g., cigarettes, alcohol) have found good support for a two-factor latent solution consisting of Amplitude and Persistence (Mackillop et al., 2009). Various teams have replicated this finding across families of drug reinforcers, including cigarettes (Bidwell et al., 2012), marijuana (Aston et al., 2017), heroin, and cocaine (Schwartz et al., 2023). Each bears close resemblance to elements either directly or indirectly estimated in the framework of Hursh & Silberberg (2008), with Amplitude most related to Q_0 (i.e., volumetric levels of consumption) and Persistence most related to α and P_{MAX} (i.e., sensitivity of consumption to changes in price). The metrics discussed here each provide information useful for evaluating reinforcer effects within and between distinct families of reinforcers. For example, a study may aim to draw comparisons between reinforcers to compare the potential of each for use and abuse (i.e., balancing therapeutic effects with potential for problematic consumption levels). Alternatively, others may aim to conduct explorations of a single drug reinforcer within in range of varying units/dosages.

Hursh & Silberberg (2008) introduced the concept of Essential Value (EV) as a strategy for supporting comparisons across drug reinforcers and across differing units (e.g., dosages). Specifically, the goal of the EV strategy was to isolate variance associated with varying units (e.g., units of reinforcers produced) and to promote a more universal interpretation of how a given reinforcer affects behavior across schedules. The expression presented by Hursh & Silberberg (2008) was later expanded and simplified in Gilroy (2023), see Eq. 2.

$$EV = P_{MAX} = \frac{1}{\alpha Q_0} * -W\left(\frac{-1}{ln10^k}\right)$$
$$Normalized\ EV = P_{MAX}Q_0 = \frac{1}{\alpha} * -W\left(\frac{-1}{ln10^k}\right) \tag{2}$$

The two expressions in Eq. 2 reflect EV with (lower) and without (upper) a normalization accounting for reinforcer unit differences. The upper expression reflects the more typical use case when patterns of consumption within or across participants are examined and share a common reinforcer unit (e.g., # of cigarettes with equal nicotine content). In contrast, the lower expression is useful in the less common case wherein patterns of consumption are simultaneously analyzed across differing reinforcer units (e.g., high-nicotine vs. low-nicotine cigarette consumption). The EV metric highlights the importance of each of the metrics discussed thus far in characterizing and comparing reinforcer scaling (i.e., Q_0 and P_{MAX}).

The metrics Q_0 and P_{MAX} provide straightforward and representative reflections of Amplitude and Persistence. That is, although parameter α can also be used to characterize Persistence (i.e., a continuous rate of change in η), parameters such as α are traditionally difficult to interpret outside of the specific contexts in which they are estimated. For example, parameter α is straightforward to estimate and compare in contexts where a common scale (k) exists but difficult to compare outside of these circumstances. In contrast, P_{MAX} reflects the scaling of reinforcers in a way that (1) is generally robust to variance in spans (ks) and (2) is easily interpreted via visual inspection of a work output curve, see Figure 1.[2] Likewise, Q_0 is also straightforward and can be easily estimated or directly sampled when the effects of pricing are absent, such as by directly observing or querying individual consumption at a P of 0 (see Amlung et al., 2015, for a relevant example of such).

Elucidating Key Metrics in Operant Demand

The most prevalent means of reporting demand intensity and the scaling of reinforcer effects take the form of fitted model parameters (i.e., Q_0, α). High levels of adoption for the Exponential model presented by Hursh & Silberberg (2008) make good sense given the high level of applicability facilitated by a small number of parameters. For example, in the absence of very small fractional consumption values (e.g., 0.001) or non-consumption (i.e., 0), the original model fares quite well with consumption values that range across multiple orders. Furthermore, metrics not revealed directly from modeling (e.g., P_{MAX}) are calculated with ease via exact solutions (Gilroy et al., 2019). Although derivatives of this framework have been put forward to accommodate cases where non-consumption values are observed (Gilroy et al., 2021; Koffarnus et al., 2015), each implementation represents a departure from the original manner of regression and presents with respective tradeoffs.

Empirical alternatives that do not require the use of parameter estimation exist for Q_0, O_{MAX}, and P_{MAX}. Demand intensity and responding representative of Q_0 can be captured simply

[2] Note. The process of estimating PMAX is complicated when parameters such as span (k) fall below an absolute minimum threshold (Gilroy et al., 2019).

by sampling consumption in the absence of cost (i.e., Q_{FREE}). Likewise, values of O_{MAX} and P_{MAX} can be inferred visually via the inspection of a total expenditure curve (e.g., Greenwald & Hursh, 2006). Although methods using empirical data are more easily performed, these present various limitations because of the simplicity of the strategy. First, researchers typically do not have a priori information regarding which prices are likely meaningful to prospective participants. Said a bit more directly, price assays for hypothetical purchase tasks are weakly informed and cover a substantial range (e.g., 0 to 10,000 USD per unit), which typically results in an oversampling of responding at higher/lower prices and undersampling of responding in the region of the curve associated with the greatest change (i.e., the bend of the curve). As a result of this lack of precision, the value of empirical data is limited because fixed pricing tasks rarely ask the *right types of questions* of the right participant. This combination of issues contributes to questionable predictions of P_{MAX} and O_{MAX} from empirical data because this approach presumes that one of the prices sampled is reasonably representative of P_{MAX} (see Gilroy et al., 2019, for representative simulations). At present, contemporary fixed pricing practices substantially limit the utility of empirical metrics in the framework (i.e., empirical P_{MAX} and demand intensity).

Adaptive Algorithms in Operant Behavioral Economics

In comparison to fixed assessments, which are identical across all respondents, adaptive assessments mutate in response to information provided by individual respondents. That is, an algorithm is put into place to present the most statistically informative questions for individual respondents. Although not yet available in work applying operant demand methods, adaptive tasks have been available for delay discounting research for some time. These include tasks that adapt to participant responses to identify a representative data point, such as 50% decay or ED50 (J. H. Yoon & Higgins, 2008), or those that identify some parameter based on an assumption of the presumed functional form of the process underlying the data (e.g., decay rate matched to 50% decay).

The task presented by Du et al. (2002) features an algorithm to iteratively elucidate a boundary wherein the value of each prospect (e.g., smaller sooner, larger later) is essentially equivalent (i.e., neither substantially better than one another) by halving the difference of the commodity up or down from the previous participant response. This is also known as an adjusting amount task, which adaptively adjusts the amount of the commodity to find a specific point (i.e., in the case of discounting, an indifference point). The goal of this process is to reveal a derived ordinate among a set of fixed delays, which yields a curve that may be characterized via statistical analysis. For the Du et al. (2002) task, the algorithm was driven by pre-defined iterative limits (i.e., a set number of adaptive choices for each delay); however, other algorithms in this space included constraints more determined by participant responses.

Johnson & Bickel (2002) used an algorithm that also adjusted the amount of the commodity based on participant responses, but how the amount was determined was based on moving upper and lower limits until the difference between the upper and lower limits was 2% or less of the larger magnitude. It warrants noting that both these tasks rely on a set of fixed delays and only the values within those delay points are assessed in an adaptive fashion (i.e., adapting amounts, not delays). Furthermore, data generated from these tasks still required nonlinear model fitting, so relevant metrics needed to be derived from the data before statistical comparisons could be performed.

Other adaptive tasks exist and free the analyst from the need to perform model fitting by providing a parameter that references a presumed data-generating process. One example is the adjusting delay task (Koffarnus & Bickel, 2014), in which a larger later option is presented alongside an immediate option half the size of the larger. Each choice changes the delay to a larger later option based on the participant's previous response throughout five questions. The last response is then used to identify the ED50, which refers to the point at which the larger, later option is subjectively equal to the smaller, sooner option and is reported as a fitted parameter, presuming a hyperbolic form (Mazur, 1987). Another example is the Three-option Adaptive Discount rate measure (ToAD; H. Yoon & Chapman, 2016), which has similar logic to the Johnson & Bickel (2002) adjusting amount task but uses three choices to shift upper and lower limits over ten choices to identify a discount rate based on a hyperbolic function. While each of these adaptive tasks avoids the need for nonlinear model fitting, they still assume a functional form for analytic purposes. Because of this, these tasks cannot be used for comparing different functional forms of the data, as they presuppose them to identify a discount rate rather than generate different subjective values to be modeled.

Despite the issues, the adaptive tasks used in discounting have still been fruitful in understanding the discounting of various commodities in different contexts. However, adapting these tasks to demand presents a challenge because of the relevant metrics used in demand (i.e., Q_0, P_{MAX}, O_{MAX}). Adaptive tasks that find points across fixed points of a factor still require model fitting, thereby adding additional considerations given the number of available demand models and the technical skill needed to explore clinically meaningful covariates (Koffarnus et al., 2022). In some cases, these individual model parameters are not meant to be interpreted on their own but instead are meant to directly identify said demand metrics (e.g., Newman & Ferrario, 2020). Adaptive tasks that identify a parameter still necessitate a functional form of the data-generating process which ignores individual variability. As such, there is presently no comparable method that corresponds to metrics and values relevant to the Operant Demand Framework.

The lack of a comparable procedure for purchase tasks in operant demand is likely due to several factors. First, these tasks do not have a ceiling or shared upper limit such as in the discounting paradigm (i.e., 100% of maximum value for a delayed/immediate choice). This invites a greater amount of variability in responding. Second, these tasks can include one or more forms of goods/services, and this adds complexity compared to simpler binary choices common in delay discounting studies (i.e., continuous vs. dichotomous responding). Third, metrics extracted from demand curves emphasize relative changes in consumption and price (e.g., P_{MAX}) and this is less straightforward than optimizing toward some fixed quantity, e.g. ED50. For these reasons, adaptive purchase tasks for operant demand likely require a more flexible approach. Thus, the creation of an adaptive demand task that identifies relevant demand metrics without the additional need for nonlinear model fitting (see Kaplan et al., 2021, for additional considerations) with a high degree of precision would decrease the barrier of entry to study demand for various commodities and simplify statistical analyses.

Machine Learning and Dynamic Adjustment Algorithms

The term Machine Learning (ML) refers to a family of methods designed to support drawing generalizable inferences from data (Blum et al., 2020). These tools are applied broadly, toward many practical and theoretical issues, and a complete review of these methods is beyond the scope of this work. Rather, the focus of this work is instead on how ML can be used to supplement contemporary methods in purchase tasks commonly used in operant demand research. The central goal of this work is to present and review a process of developing an agent that responds to the input of a user and guides the presentation of future pricing questions as a function of informational value (i.e., associated with greater resource allocation/responding).

Reinforcement Learning (RL) can be viewed as a derivative of ML; however, RL is distinct from both supervised and unsupervised forms of ML. For instance, ML is traditionally applied to either extract structure from data or perform classification, whereas RL is often used to guide the making of "sequential optimal decisions under Uncertainty" via a Markov Decision Process (Rao & Jelvis, 2023). These tools are frequently used to model decision-making processes (e.g., an adversary for computer games) through the design of agents that suggest actions (K), given prior/available information and environmental state. The choice to demonstrate a given action (k) is conditioned on a history of prior reinforcement and present conditions. The reinforcement element of this approach refers to a weighting of respective actions (K) given the history and likely future of rewards. Various algorithms that differentially weight actions differ in how each balances the need for exploring available actions and for exploiting prior experience.

Actions available to agents are favored or made more likely based on the concept of *regret*. Regret refers to the evaluation of observed rewards associated with actions (i.e., more regret equates to exploring an underperforming action). The selection of actions by the agent is driven by levels of regret, whereby the most likely action to pursue is the one that maximizes the probability and quantity of reward (i.e., minimizing regret). This general process can be adjusted to balance the need for exploring unexplored actions and exploiting prior knowledge to seek optimal rewards.

Algorithm 1: Partially Ordered Set Master Algorithm

Require: Partial order ($\preceq$) for K different settings (ks), reweighting constant $\beta \in (0,1)$, and agent observations across time (O_t)

$\forall k \in K$: let $w(k) = 1$

for t = *1, 2...* **do**

$\forall k \in K$: let $A_t(k) = \sum_{x \in K, x \succeq k} w_t(x)$

$\forall k \in K$: let $B_t(k) = \sum_{x \in K, x \preceq k} w_t(x)$

Predict $k_t = argmax_{k \in K}\ min\{A_t(k), B_t(k)\}$

Observe $o_t \in \{-1, 0, 1\}$

if $o_t = +1$ **then**

$$\forall k \in K\text{: let } w_{t+1}(k) = \begin{cases} \beta w_t(k), & k \preceq k_t \\ w_t(k), & \text{otherwise} \end{cases}$$

if $o_t = -1$ **then**

$$\forall k \in K\text{: let } w_{t+1}(k) = \begin{cases} \beta w_t(k), & k \succeq k_t \\ w_t(k), & \text{otherwise} \end{cases}$$

The choice of algorithms incorporated in RL approaches is guided based on various assumptions for the decision-making process and the types of data being optimized (Rao & Jelvis, 2023). For example, there are often distributional assumptions regarding the probability or magnitude of reward for specific actions (e.g., Binomial for yes [1] or no [0] responses). Additionally, the relative superiority of an action may not be stationary, and different actions may have superior outcomes at different points in time. Algorithms for dynamically predicting optimal actions vary considerably and are carefully selected or designed on the specific nature of the task, context, and manner of reward.

Partially Ordered Set Master Algorithm

The Partially Ordered Set Master (POSM) algorithm is a variant of RL that explores various actions that are ordered (Missura & Gärtner, 2011). For example, the ordering may correspond to multiple settings that vary in terms of increasing difficulty to a hypothetical user.

The POSM algorithm is unique in that the available actions (i.e., $k \in K$) are ordered rather than each action having a discrete distribution. The ordering inherent in the POSM algorithm is useful in purchase tasks because ordering naturally exists among pricing options and because there is no assumption that the price associated with peak expenditure is stationary over time. The algorithm presented in Missura & Gärtner (2011) is depicted in Algorithm 1.

The original purpose of the algorithm was to assist in identifying an optimal setting regarding user performance (i.e., neither too easy nor too hard). The use of the term 'master' is related to the game-based context and the design of an intelligent adversary (i.e., as if opposing a ‘master' in some game-like context). Beliefs regarding specific settings are updated using performances demonstrated by the user, particularly when a given setting appears 'too easy' (-1) or 'too hard' (+1) for the user. The total mass of beliefs is reflected across policies for 'too low' and 'too high' at a given time (*t*), represented for each action (*k*) as A_t and B_t, respectively. Each of these are listed below in Eq. 3.

$$A_t(k) = \sum_{x \in K, x \geqslant K} w_t(x)$$
$$B_t(k) = \sum_{x \in K, x \leqslant K} w_t(x) \tag{3}$$

The process of re-weighting beliefs across prices according to a fixed constant is illustrated in Algorithm 1. This approach is efficient in terms of maximizing information because the process of updating beliefs, $\beta w_t(x)$, carries forward to all levels above or below action k. Estimates of beliefs favoring specific actions at a specific time (θ_t) are determined using the minimum of both policies for A and B for each level of k at a given time *t*. The specific calculation for this is provided below in Eq. 4.

$$\hat{\theta}_t = argmax\ min\{A_t(\hat{\theta}_t), B_t(\hat{\theta}_t)\} \quad (4)$$

An Algorithm-driven Hypothetical Purchase Task

The POSM algorithm can be adapted for use in purchase to explore levels of price (*P*) using the total amount of expenditure (i.e., *PQ*) and the concept of regret across *K* levels of *P*. The minimization of regret dynamically guides the exploration of prices in search of a point of peak expenditure. The point at which peak expenditure is optimized reveals the empirical P_{MAX} and O_{MAX} for the user. A visual of the link between the demand curve, the point of unit elasticity (P_{MAX}), and the point of peak work (O_{MAX}) is illustrated in Figure 1.

Figure 2 provides an overview of how the POSM algorithm can be used to guide questions presented to users in the context of a purchasing task. The begins with the initialization of a vector of beliefs at *K* levels of *P* (e.g., 1-500 USD; *w* = 1) and the generation of an initial prediction. Per Eq. 4, this is essentially a uniform prior, and the initial prediction is the midpoint of the range of levels (e.g., *P* of 250 USD for pricing from 1 to 500 USD). Input from the user across iterations reveals expenditure (*R*) at respective levels of price (*k*) and beliefs are updated to favor levels that minimize regret (i.e., produce value closest to presently known O_{MAX}). This process is either repeated for a fixed number of iterations (*t*) or terminated once a threshold is met and further iterations are unlikely to further reduce regret. Figure 3 illustrates a simulated sequence wherein the agent adapts to the expenditure of the simulated user and guides the prices presented to more sample levels at or near P_{MAX} (see left). The overall regret decreases as the user provides information that more consistently produces the highest expenditure (O_{MAX}; see right). The empirical data comprising the work output and demand functions learned from the simulated user in the task are presented in Figure 4.

Both Figure 3 and Figure 4 illustrate how a reinforcement learning approach pairs quite elegantly with methods designed to explore reinforcer value. The POSM algorithm is quite suited given the strategic use of ordinal information and usefulness in seeking questions that provide the most informational value (i.e., nearer P_{MAX}) and avoiding values that are seldom useful (i.e., non-consumption at prohibitively expensive prices).

Research Questions

The primary goal of this research was to present and evaluate an adaptive approach to evaluating reinforcer value for use in adaptive purchasing tasks. This work primarily focused on whether the algorithm, across varying question lengths, could reliably and efficiently recover unknown values of P_{MAX}. This algorithm was evaluated in two dimensions with research questions specific to each.

Research Question 1 (RQ1): Given simulated P_{MAX} values, does the POSM algorithm recover the price associated with peak work (P_{MAX}) in purchase tasks consisting of 5, 10, 15, and 20 sequential questions?

Research Question 2 (RQ2): Related to RQ1, what is the minimum number of sequential questions necessary to recover statistically equivalent estimates of P_{MAX}?

Method

Simulated Agents

A total of 4,000 (n = 1,000 x 4 run lengths) simulated agents were generated across four different hypothetical purchase task lengths: 5, 10, 15, and 20 questions. Each simulated series sampled a randomly selected O_{MAX}, P_{MAX}, and Q_0 from the ranges of 50-950, 1,000-5,000, and 10-100, respectively. The quantity expended by the hypothetical user was generated by using such values and the solutions provided in Gilroy (2023) and Gilroy et al. (2019) to produce a prediction using the Exponential model proposed by Hursh & Silberberg (2008). All simulations were conducted using the R Statistical Program (R Core Team, 2013), and the same seed value was used across simulations to isolate differences solely due to iteration length.

Analytic Strategy

Four individual equivalence tests were conducted for each of the varying numbers of questions in the task. All equivalence tests were performed using the TOSTER R package (Lakens, 2017; Lakens & Caldwell, 2022). For each test, the smallest effect size of interest was for differences between simulated and true P_{MAX} values was set to a value of 0.01. The 0.01 value on the log scale provides a convenient means of approximating an estimated 1% difference between paired values and differences below this upper and lower threshold were not considered to be statistically meaningful.

Results

RQ1. Tests of Equivalence for P_{MAX}

Illustrations of task equivalence are illustrated in the righthand portion of Figure 5 for tasks featuring 5, 10, 15, and 20 sequential questions. Task performance with 5 questions was not statistically different (i.e., interval did not include 0) but the two did not demonstrate statistical equivalence (i.e., interval exceeded SESOI bounds). The performance of tasks that included 10, 15, and 20 questions were not significantly different and demonstrated statistical equivalence.

RQ2. Effects of Iteration Length on P_{MAX}

Visualizations of task correspondence across varying task lengths are provided in the lefthand portion of Figure 5. Visual inspection revealed strong overall correspondence between estimates of P_{MAX} across task lengths. Tasks that featured 5 sequential questions yielded P_{MAX} values that were highly correlated with seed P_{MAX} values ($r = 0.97$, $p = 0$); however, tasks at this length did not demonstrate statistical equivalence. All remaining task lengths demonstrated essentially perfect correspondence, with lengths of 10 ($r = >.999$, $p < .001$), 15 ($r = >.999$, $p = .001$), and 20 ($r = >.999$, $p = .001$) questions producing statistically equivalent estimates of P_{MAX}.

Discussion

Purchase tasks are among the most frequently used tools in research applying the Operant Demand Framework. Data from these tasks are readily analyzed using any of the modeling options derived from the framework of Hursh & Silberberg (2008). However, despite good adoption and flexibility, the fixed and standardized nature of pricing assays included in these tasks limits research on individual decision-making in several regards. For example, the fixed nature of contemporary pricing assays limits the usefulness of empirical metrics of EV because the price points sampled seldom closely correspond with the point at which demand for a reinforcer shifts from inelastic to elastic (e.g., P_{MAX}, O_{MAX}). The goal of this work was to provide an introduction and evaluation of an adaptive approach based on RL for purchase tasks that could be used to extend research in operant demand.

The results from this study revealed strong overall performance across tasks of all lengths and good correspondence was observed between seeded P_{MAX}, even with the most abbreviated forms of the task (i.e., just five questions in a given task). However, despite good correspondence, differential degrees of statistical equivalence were observed for tasks across varying batches of questions presented to simulated agents. The most abbreviated form of the task, which entailed a series of 5 questions, was significantly correlated with a seeded P_{MAX}, but the overall evaluation did not support a determination of statistical equivalence. In contrast, all evaluations with lengths of ten or more questions demonstrated statistical equivalence and consistently recovered the seeded values.

Overall findings here suggest that more abbreviated adaptive tasks may be less capable of capturing P_{MAX} values in several cases, and this warrants a bit of discussion. First, shorter tasks may be less reliable when the range of prices considered is substantial (e.g., 0-1000 USD/unit) and/or the participant's P_{MAX} value falls near the extremes (e.g., 950 USD/unit). Given that the task operates from an initial uniform prior, larger ranges of questions/updates would be necessary to ensure enough freedom for the algorithm to explore the parameter space nearer the extremes (i.e., larger ranges = more iterations required). Overall, providing a larger series of questions to the user would provide the algorithm with additional opportunities to sufficiently recover from a chance errant response from the user (i.e., user error). This feature of adaptive purchase tasks is useful because a sufficient number of questions asked in this manner may lessen the risk of certain data being determined 'unsystematic' and not amenable to statistical analysis.

Simulations performed in this study revealed that adaptive purchase tasks with ten or more sequential questions were essentially perfect in providing empirical measures of P_{MAX} that were identical to seeded P_{MAX} values. This finding suggests that RL algorithms could enhance the flexibility and utility of purchase tasks and may obviate the need for more statistically and mathematically complex operations when using fixed price arrangements. Specifically, nonlinear models, and the solutions necessary for deriving key metrics of essential value from them, may not be necessary to extract values of critical interest within the current framework. That is, both Q_0 (i.e., sampling consumption at P of 0) can P_{MAX}/O_{MAX} can be revealed with good consistency via empirical means and without the need for statistical determination.

Implications for Modelling in Operant Demand

The framework of Hursh & Silberberg (2008) has been critical in guiding modern approaches for evaluating choice behavior under constraints (e.g., varying prices, availability of alternatives). The Exponential model of operant demand and its derivatives have provided a reliable means of elucidating critical elements in the Operant Demand Framework (e.g., revealing Q_0 and P_{MAX}). Modeling has traditionally been necessary because fixed pricing assays seldom provided information sensitive enough to support comparisons within or between participants. However, given that pricing assays can be made adaptive and directly reveal and sample relevant prices, nonlinear modeling such as that suggested by Hursh & Silberberg (2008) may not be required to answer many common research questions.

Although there is an active community of researchers with statistical training working to provide guidelines and support for the extension of nonlinear modeling practices, findings from this study prompt questions regarding the necessity of nonlinear methods in the framework. That is, researchers may not need to extend beyond simpler tests (e.g., T-tests, ANOVA) to make group-level comparisons based on specific empirical values (e.g., Q_0). The tools presented in this work suggest that newer adaptive tasks could address prior shortcomings without reliance on nonlinear methods and a departure from a modeling-forward approach has the potential to simplify current practices in several ways.

First, the strategy provided here avoids making assumptions about the underlying processes involved in decision-making, which is an assumption that is not easily resolved. The adaptive task presented in this work is not bound to a specific underlying process, which may vary across individuals/organisms, and does not require an opinion regarding which model form is most suited to all in a study. Given that the true data-generating process is unknown, the strategy described here provides a method for deriving metrics relevant to EV that avoids many problematic assumptions.

Second, related to the prior point, fixed pricing arrangements oversample consumption at pricing extremes and this reveals substantial variability in the ranges of consumption observed for individuals and groups. Most models derived from the framework of Hursh & Silberberg

(2008) assume a common range of consumption between pricing extremes and issues regarding the modeling of zero and non-zero lower limits remain unresolved at this time (Gilroy, 2022; Gilroy et al., 2021; Koffarnus et al., 2015). The approach presented here provides a path forward for this issue by focusing less on price extremes and instead on prices within the most information portion of the demand curve, those near P_{MAX} and the point of unit elasticity. This approach provides a means of extracting key metrics of operant demand without the need to grant unnecessary influence to consumption at pricing extremes (e.g., consumption values at or near zero) and make strong assumptions regarding a range of consumption shared across all consumers in a group or sample. This strategy provides a parsimonious alternative to contemporary operant demand methods and addresses historical issues in the framework within the task presented rather than the modeling performed.

Third, and lastly, addressing known issues with fixed purchase tasks by using an adaptive algorithm does not limit any actions available to the analyst. Specifically, the data gathered from adaptive purchase tasks reveals empirical metrics with high precision and also provides empirical choice data that remains amenable to nonlinear statistical analysis. For example, the analyst may use such tasks to extract key metrics of demand empirically but also use individual-level responses to characterize the nature and form of those curves. Relatedly, parameters such as α are trivial to solve without regression when information regarding Q_0 and P_{MAX} are already known (see Appendix).

Limitations and Areas of Future Extension

The methods presented in this work provide an encouraging extension of RL methods that are conceptually consistent with the Operant Demand Framework and provide a potential alternative to unnecessarily complex nonlinear modeling. The results presented here are highly encouraging, suggesting that nonlinear modeling may not be a prerequisite to the application of the Operant Demand Framework and that many long-standing issues with nonlinear models may be avoided altogether. However, despite encouraging findings, further evaluation with real-world applications is necessary to better understand how this novel approach can best support work in Operant Behavioral Economics. Additionally, statistical training and expertise in computer science are not skills that are reflected in the operant training tradition, and the implications of increased reliance on algorithms in applied research are not yet fully understood.

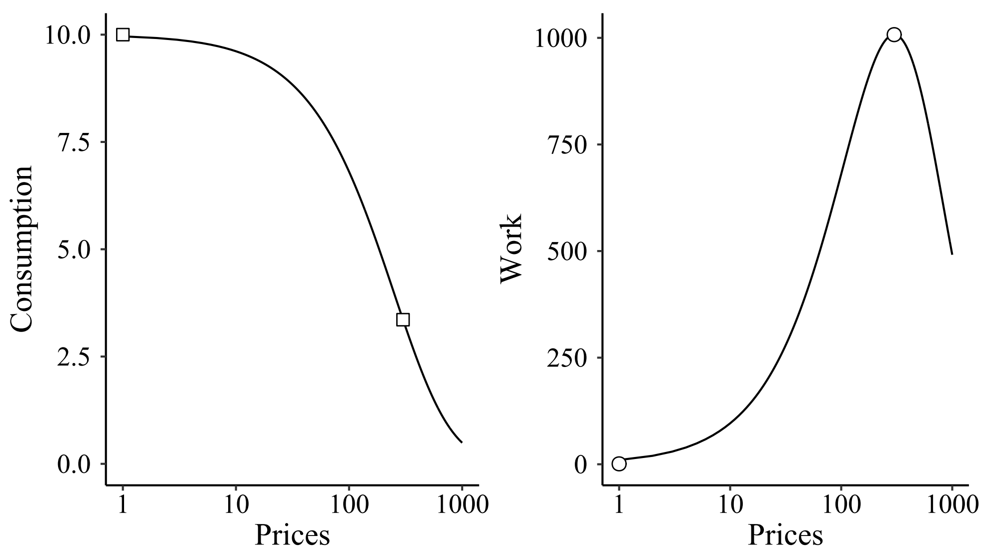
Appendix

Extracting Model Values from Model-free Estimates

The quantity P_{MAX} equates to α given respective units. The results of empirical fitting yielding Q_0 and P_{MAX} can be used to solve for α (given any suitable value for *k*). Relevant calculations are provided below illustrating respective calculations (left) and a worked solution (right). The plot below illustrates the range of prices with the solved *α* parameter.

$$P_{MAX} = \frac{1}{\alpha Q_0} * -W(\frac{-1}{ln10^k})$$
$$\alpha = \frac{-W(\frac{-1}{ln10^k})}{P_{MAX}Q_0}$$

$$Q_0 = 10$$
$$P_{MAX} = 300$$
$$k = 3$$
$$\alpha = \frac{-W(\frac{-1}{ln10^k})}{P_{MAX}Q_0}$$
$$5.73e^{-05} = \frac{-W(\frac{-1}{ln10^3})}{300 * 10}$$



References

Amlung, M., McCarty, K. N., Morris, D. H., Tsai, C.-L., & McCarthy, D. M. (2015). Increased behavioral economic demand and craving for alcohol following a laboratory alcohol challenge. *Addiction*, *110*(9), 1421–1428. https://doi.org/10.1111/add.12897

Aston, E. R., Farris, S. G., MacKillop, J., & Metrik, J. (2017). Latent factor structure of a behavioral economic marijuana demand curve. *Psychopharmacology*, *234*(16), 2421–2429. https://doi.org/10.1007/s00213-017-4633-6

Bidwell, L. C., MacKillop, J., Murphy, J. G., Tidey, J. W., & Colby, S. M. (2012). Latent factor structure of a behavioral economic cigarette demand curve in adolescent smokers. *Addictive Behaviors*, *37*(11), 1257–1263. https://doi.org/10.1016/j.addbeh.2012.06.009

Blum, A., Hopcroft, J., & Ravindran, K. (2020). *Foundations of Data Science*. Cambridge University Press.

Du, W., Green, L., & Myerson, J. (2002). Cross-cultural comparisons of discounting delayed and probabilistic rewards. *The Psychological Record*, *52*(4), 479–492. https://doi.org/doi.org/10.1007/BF03395199

Gilroy, S. P. (2022). Hidden equivalence in the operant demand framework: A review and evaluation of multiple methods for evaluating nonconsumption. *Journal of the Experimental Analysis of Behavior*, *117*(1), 105–119. https://doi.org/doi.org/10.1002/jeab.724

Gilroy, S. P. (2023). Interpretation(s) of essential value in operant demand. *Journal of the Experimental Analysis of Behavior*, *119*(3), 554–564. https://doi.org/10.1002/jeab.845

Gilroy, S. P., Kaplan, B. A., Reed, D. D., Hantula, D. A., & Hursh, S. R. (2019). An exact solution for unit elasticity in the exponential model of operant demand. *Experimental and Clinical Psychopharmacology*, *27*(6), 588–597. https://doi.org/10.1037/pha0000268

Gilroy, S. P., Kaplan, B. A., Schwartz, L. P., Reed, D. D., & Hursh, S. R. (2021). A zero-bounded model of operant demand. *Journal of the Experimental Analysis of Behavior*, *115*(3), 729–746. https://doi.org/10.1002/jeab.679

Greenwald, M. K., & Hursh, S. R. (2006). Behavioral economic analysis of opioid consumption in heroin-dependent individuals: Effects of unit price and pre-session drug supply. *Drug and Alcohol Dependence*, *85*(1), 35–48. https://doi.org/10.1016/j.drugalcdep.2006.03.007

Hursh, S. R. (1980). Economic concepts for the analysis of behavior. *Journal of the Experimental Analysis of Behavior*, *34*(2), 219–238. https://doi.org/doi.org/10.1901/jeab.1980.34-219

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

Hursh, S. R., & Bauman, R. A. (1987). The behavioral analysis of demand. In L. Green & J. Kagel (Eds.), *Advances in Behavioral Economics* (Vol. 1, pp. 117–165). Ablex Publishing Corporation.

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

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, *77*(2), 129–146. https://doi.org/10.1901/jeab.2002.77-129

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*(2–3), 333–358. https://doi.org/10.1007/s40614-021-00299-7

Koffarnus, M. N., & Bickel, W. K. (2014). A 5-trial adjusting delay discounting task: Accurate discount rates in less than one minute. *Experimental and Clinical Psychopharmacology*, *22*(3), 222–228. https://doi.org/10.1037/a0035973

Koffarnus, M. N., Franck, C. T., Stein, J. S., & Bickel, W. K. (2015). A modified exponential behavioral economic demand model to better describe consumption data. *Experimental and Clinical Psychopharmacology*, *23*(6), 504–512. https://doi.org/10.1037/pha0000045

Koffarnus, M. N., Kaplan, B. A., Franck, C. T., Rzeszutek, M. J., & Traxler, H. K. (2022). Behavioral economic demand modeling chronology, complexities, and considerations: Much ado about zeros. *Behavioural Processes*, *199*, 104646. https://doi.org/10.1016/j.beproc.2022.104646

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, *8*(4), 355–362. https://doi.org/doi.org/10.1177/1948550617697177

Lakens, D., & Caldwell, A. (2022). *Package 'TOSTER'* [Computer software].

Mackillop, J., Murphy, J. G., Tidey, J. W., Kahler, C. W., Ray, L. A., & Bickel, W. K. (2009). Latent structure of facets of alcohol reinforcement from a behavioral economic demand curve. *Psychopharmacology*, *203*(1), 33–40. https://doi.org/10.1007/s00213-008-1367-5

Madden, G. J., Smethells, J. R., Ewan, E. E., & Hursh, S. R. (2007). Tests of behavioral-economic assessments of relative reinforcer efficacy: Economic substitutes. *Journal of* *the Experimental Analysis of Behavior*, *87*(2), 219–240. https://doi.org/doi.org/10.1901/jeab.2007.88-355

Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. In *The effect of delay and of intervening events on reinforcement value* (pp. 55–73). Lawrence Erlbaum Associates, Inc.

Missura, O., & Gärtner, T. (2011). Predicting Dynamic Difficulty. *Advances in Neural Information Processing Systems*, *24*. https://papers.nips.cc/paper_files/paper/2011/hash/7c9d0b1f96aebd7b5eca8c3edaa19ebb-Abstract.html

Newman, M., & Ferrario, C. R. (2020). An improved demand curve for analysis of food or drug consumption in behavioral experiments. *Psychopharmacology*, *237*(4), 943–955. https://doi.org/10.1007/s00213-020-05491-2

R Core Team, R. (2013). *R: A language and environment for statistical computing*.

Rao, A., & Jelvis, T. (2023). *Foundations of reinforcement learning with applications in finance* (1 Edition). Chapman & Hall, CRC Press.

Schwartz, L. P., Toegel, F., Devine, J. K., Holtyn, A. F., Roma, P. G., & Hursh, S. R. (2023). Latent factor structure of behavioral economic heroin and cocaine demand curves. *Experimental and Clinical Psychopharmacology*, *31*(2), 378–385. https://doi.org/10.1037/pha0000594

Strickland, J. C., Reed, D. D., Hursh, S. R., Schwartz, L. P., Foster, R. N. S., Gelino, B. W., LeComte, R. S., Oda, F. S., Salzer, A. R., Schneider, T. D., Dayton, L., Latkin, C., & Johnson, M. W. (2022). Behavioral economic methods to inform infectious disease response: Prevention, testing, and vaccination in the COVID-19 pandemic. *PLoS ONE*, *17*(1), e0258828. https://doi.org/10.1371/journal.pone.0258828

Yoon, H., & Chapman, G. B. (2016). A Closer Look at the Yardstick: A New Discount Rate Measure with Precision and Range. *Journal of Behavioral Decision Making*, *29*(5), 470–480. https://doi.org/10.1002/bdm.1890

Yoon, J. H., & Higgins, S. T. (2008). Turning k on its Head: Comments on use of an ED50 in Delay Discounting Research. *Drug and Alcohol Dependence*, *95*(1–2), 169–172. https://doi.org/10.1016/j.drugalcdep.2007.12.011

Figure 1. Prototypical Demand and Revenue Functions

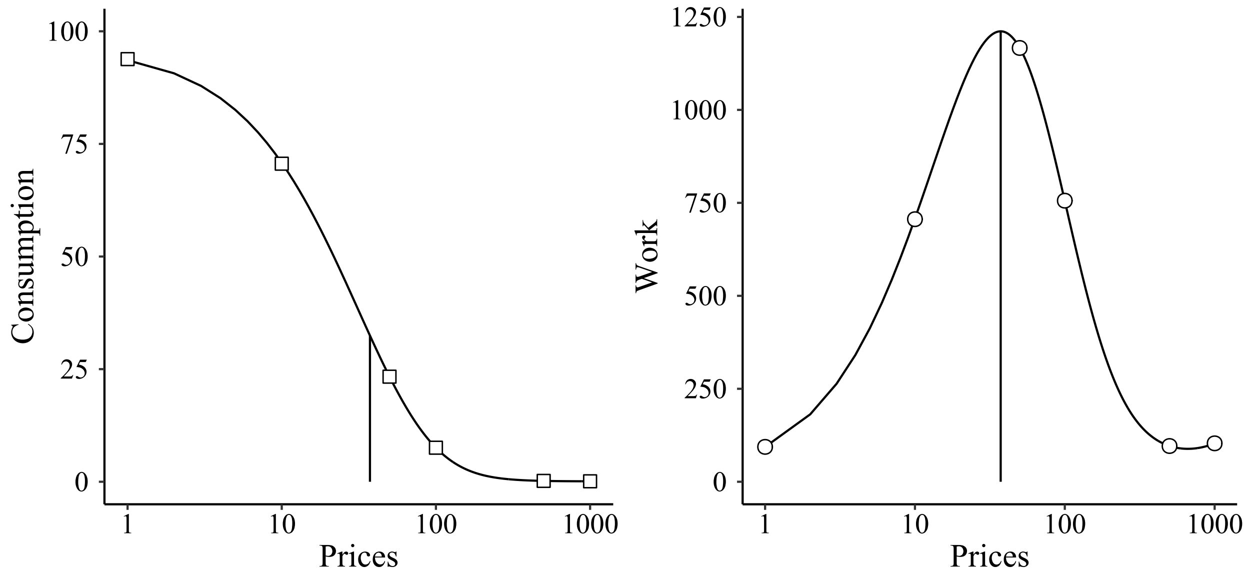


Figure 2. Workflow for Adaptive Hypothetical Purchase Task

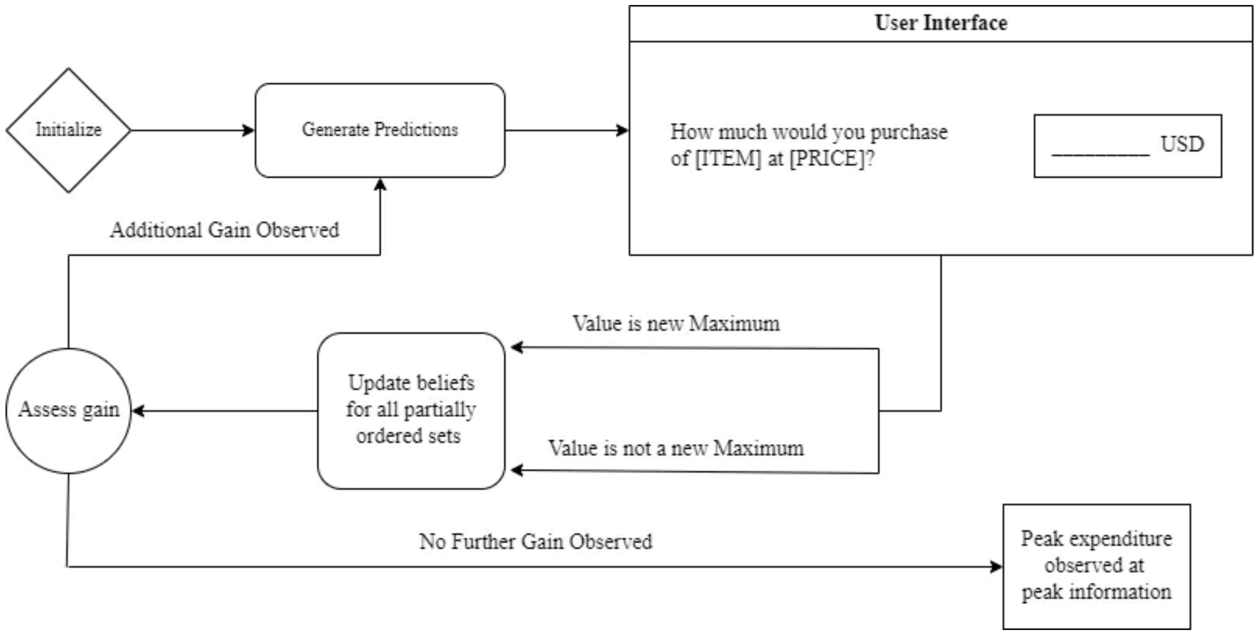


Figure 3. Price Beliefs, Expenditure, and Regret Minimization

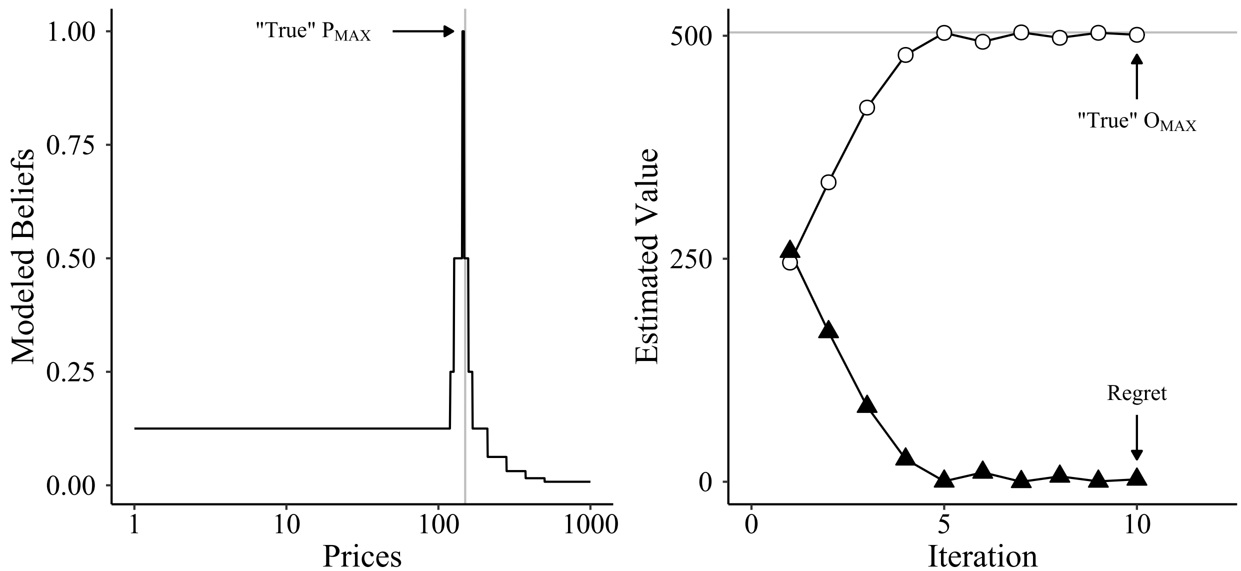


Figure 4. Price and Consumption Sampling from Adaptive Task

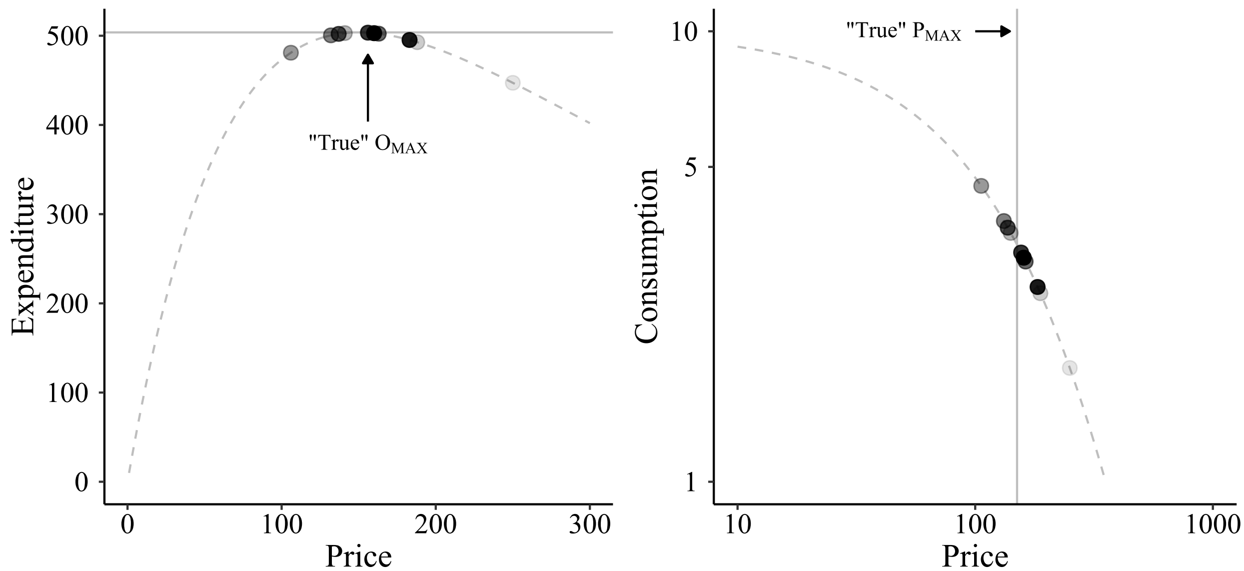


Figure 5. Correspondence and Equivalence for P_{MAX} Values