Identification and Estimation of Forward-looking Behavior: The Case of Consumer Stockpiling*

Andrew T. Ching† and Matthew Osborne‡

First draft: April 13, 2015

This draft: October 17, 2018

Abstract

This paper quantifies the strength of forward-looking behavior and its implications for pricing strategy in the context of a dynamic consumer stockpiling model. In prior work, researchers have assumed that storage costs are a continuous function of inventory. We demonstrate that this seemingly innocuous simplifying assumption rules out exclusion restrictions which naturally arise from the institutional features of the stockpiling problem. The lack of exclusion restrictions in earlier models requires researchers to fix the discount factor, instead of estimating it. We show formally that by properly modelling storage cost as a step function of inventory (because storage cost depends on the number of packages stored, instead of the actual amount of inventory), the key state variable of this model, inventory, provides *natural exclusion restrictions* that can help identify the model's parameters, including the discount factor. We then estimate a stockpiling model using scanner data on laundry detergents, and recover the population distribution of the discount factor. Our estimates suggest that although some consumers have discount factors close to the rational expectations benchmark of about 0.999, most are much less forward-looking: The estimated average weekly discount factor is about 0.71. A counterfactual exercise shows that if one used a model which fixed the discount factor to be consistent with rational expectations, one would overpredict the effect of increased promotional depth on quantity sold by 15%.

Key words: Discount Factor, Exclusion Restriction, Stockpiling, Dynamic Programming

*We thank Øystein Daljord, Aviv Nevo, Stephan Seiler and seminar participants at the Summer Institute for Competitive Strategy, the Stanford Institute for Theoretical Economics, Boston College, McMaster University, University of Toronto, University of Western Ontario, The Wharton School, Purdue University, McGill University, University of Hong Kong, Fudan University, and University of Rochester for providing helpful feedback on earlier drafts.

†University of Toronto, Rotman School of Management. email: andrew.ching@rotman.utoronto.ca.

‡University of Toronto, Institute for Management and Innovation and Rotman School of Management. email: matthew.osborne@rotman.utoronto.ca

1 Introduction

Forward-looking behavior is a critical component of many quantitative models of consumer behavior used by researchers in marketing and economics (Erdem and Keane (1996), Crawford and Shum (2005), Erdem, Imai, and Keane (2003), Hendel and Nevo (2006a), Hartmann and Nair (2010), Chan, Narasimhan, and Zhang (2008), Seiler (2013), Liu and Balachander (2014), Osborne (2011), Yao, Mela, Chiang, and Chen (2012), Yang and Ching (2014), Osborne (2018a)). When consumers are forward-looking, they also behave strategically when making their purchase decisions. The presence of strategic behavior has important implications for the profitability of a firm's dynamic pricing strategy. For example, in the context of storable consumer packaged goods such as canned tuna or canned soup, forward-looking consumers may respond to a temporary price promotion today by stockpiling the product, since they understand that future prices are likely to be high (Erdem, Imai, and Keane (2003), Liu and Balachander (2014), Haviv (2014)). If all shoppers were extremely forward-looking and acted in such a savvy way, durable goods producers would not be able to use a price skimming strategy (Coase 1972), and grocery stores or supermarkets would never sell their carried items at a regular (non-deal) price. This is not the case, since in reality price skimming and Hi-Lo pricing are both prevalent, and consumers do make purchases of products when their prices are high. Prior research on periodic promotions (Hendel and Nevo (2013), Hong, McAfee, and Nayyar (2002), Pesendorfer (2002), Sobel (1984)) has recognized that firms can use periodic promotions to price discriminate between patient and impatient consumers. Forward-looking behavior is an important driver of choice in many other marketing contexts as well: For example, for new durable goods such as cameras or smart phones, consumers may wait to purchase a product if they expect the product's price to fall in the future.

A widely used modeling framework for empirically analyzing forward-looking decisions is the discrete choice dynamic programming framework (Rust 1994). In dynamic discrete choice models, the strength of forward-looking behavior is captured by a parameter called a *discount factor*: the closer the discount factor is to 1, the more weight consumers put on future payoffs when making current decisions. Most empirical research in this area does not estimate the discount factor, instead exercising a "rational expectations" assumption which uses the prevailing interest rate to fix the discount factor accordingly. This calibration approach leads to a value of the yearly discount factor of 0.95, and the weekly discount factor of about 0.9995.[1] The rational expectations assumption

[1] At a yearly interest rate of 5%, which is consistent with U.S. real interest rates in the period 2001-2010, a rational consumer would discount utility in the following year at a rate of about 0.95, and would have a weekly discount rate of about $(1/(1+0.05)^{(1/52)} \approx 0.9995$. In practice, researchers will sometimes set a slightly lower discount factor than the rational expectations assumption implies to reduce the computational burden of estimating the dynamic model.

is, however, at odds with a wealth of experimental studies which find that the discount factors can range from between 0.00 to 0.99 (Frederick, Loewenstein, and O'Donoghue (2002)). Such a wide range of estimates suggests that the discount factor could be context specific. Moreover, in stated choice experiments performed by Dubé, Hitsch, and Jindal (2014), consumers appear to be much less forward-looking than economic theory implies, with average discount rates of 0.43. Dubé, Hitsch, and Jindal (2014) also find substantial heterogeneity in discount factors across individuals.

The reason why the discount factor is typically not estimated in discrete choice dynamic programming problems stems from an identification problem: If the researcher does not impose any functional form assumptions on the structure of the current period utility function, the discount factor cannot be identified (see Rust (1994), Magnac and Thesmar (2002)). Researchers use this as a reason for fixing the discount factor using the prevailing interest rate, instead of estimating it. However, if the true discount factor is significantly different from the one calibrated by the interest rate, the rest of estimated structural parameters will likely be biased, and so will corresponding counterfactual policy predictions.

Recently, Magnac and Thesmar (2002) and Fang and Wang (2015) have argued that if a dynamic model has exclusion restrictions, then the discount factor can be identified. Roughly speaking, exclusion restrictions occur when there exists at least one state variable that impacts a consumer's future payoffs, but not her current payoffs. The intuition behind how this leads to identification of the discount factor is that if a consumer is completely myopic, then the consumer's choice should be independent of that variable. The extent to which a consumer's choice is influenced by the state variable when the exclusion restrictions hold provides information about how forward-looking the consumer is. However, despite this recent development, most recent research that models forward-looking consumers still fixes the discount factor according to the interest rate. This is because the state variables of these dynamic models do not provide exclusion restrictions.

Our paper contributes to the literature by arguing that in the context of consumer stockpiling, one of the key state variables, inventory, provides *natural exclusion restrictions* if the storage cost is modelled properly.[2] Our key insight is that in many product categories, a consumer's current payoff depends on the storage cost of inventory, and the storage cost is not affected by changes in inventory unless a package of a product runs out. As a result, for most inventory levels a consumer's current payoff does not vary with inventory. To illustrate this insight, consider an example drawn from the laundry detergent market. Suppose a consumer has a single bottle of laundry detergent

For example, Seiler (2013) uses a value of 0.998.

[2] By "natural exclusion restrictions," we mean the exclusion restrictions are well-justified by the institutional details of the environment being studied.

in her home. This consumer has a habit of wearing clean clothes every day, and as a result does a single load of laundry every week. If the consumer is forward-looking, as she keeps consuming the laundry detergent, she may worry that if she does not buy another bottle soon when the price is low, she may be forced to buy it at a higher regular price when she uses it up in the near future. This sense of urgency will become stronger as inventory (i.e., the amount of detergent in the bottle) runs down, and her demand would appear to become more sensitive to price cuts. Moreover, for any amount of inventory remaining, the more forward-looking a consumer is, the more intense this feeling of urgency will get.

We establish that the above intuition can be formalized by developing a proof that under relatively mild conditions, all the parameters of a stockpiling model, and in particular the discount factor, can be uniquely identified. This is a useful contribution to the literature on identification, because Abbring and Daljord (2018) recently have shown that even when a dynamic model has an exclusion restriction, the discount factor may not be point identified. The intuition behind our identification proof is that the discount factor is increasing in the rate at which an individual's purchase probability changes with inventory at high levels of inventory, relative to low levels of inventory, for levels of inventory where storage costs are fixed (where the exclusion restrictions hold). For a relatively myopic individual, her purchase probability will only start to increase substantially when she gets very close to running out, resulting in an estimated discount factor that is small. The rest of the model parameters (i.e., stockout costs and storage costs) can then be derived from the equations which define choice probabilities at different inventory states. We should highlight that in the previous structural stockpiling models, researchers assume that storage costs increase continuously with inventory. However, this seemingly innocuous simplifying assumption rules out our identification arguments. If the storage cost function is convex and continuous, even a myopic individual's purchase probability will increase smoothly and at an increasing rate as inventory drops, making identification of forward-looking behavior difficult.

We demonstrate the managerial relevance of estimating the discount factor with an empirical exercise, where we estimate a dynamic structural model of consumer stockpiling on scanner data for laundry detergents, both allowing the discount factor to be free, and fixing it to the rational expectations benchmark. We then construct a counterfactual pricing scenario and show that the expected revenues from changing the pricing process for a popular detergent brand can be sensitive to the discount factor. As far as we know, our paper is the first to empirically recover consumer discount factors in field data for a stockpiling problem. We additionally allow for individual-level unobserved heterogeneity in the discount factor, and find that consumers differ substantially in their myopia. About one-quarter of the population has discount factors that are close to 1, which is consistent with the rational expectations benchmark. The rest of the population is much less forward-looking, and has discount factors that are spread between 1 and about 0.05. In particular, the average discount factor is 0.71, and the first tertile is 0.62, which are both much less than the rational expectations benchmark. Most of this heterogeneity seems to be driven by unobserved factors, a finding that is also consistent with Dubé, Hitsch, and Jindal (2014).

Turning to our pricing counterfactual, we compare the impact of changing the the pricing policy of a popular laundry detergent product on quantity sold and revenue, under the model with estimated discount factors, and the model that imposes the rational expectations assumption. In the laundry detergent category, many stores use a hi-lo pricing strategy, where prices stay fixed at a regular retail price for most of the time, but for short periods dip to a low price of about 25% less than the regular retail price. Periodic promotions such as this are designed to encourage stockpiling, as forward-looking consumers will time their purchases to occur when promotions happen. In our counterfactual pricing exercise, we increase the depth of promotions, and find that the rational expectations model specification predicts an increase in both quantity sold and revenue due to the policy that is about 15% higher than the forecast from the model specification where the discount factor is estimated. To see the intuition behind this counterfactual result, suppose that individuals are very forward-looking. As one increases the attractiveness of promotions for a particular product, forward-looking individuals will be more likely to wait for promotions to occur on that product, as opposed to purchasing something else, as their inventory drops. One will overstate the strength of this effect the more forward-looking one assumes individuals are. We note that we also examine the impact of increasing promotional frequency, but find that this pricing policy is overall less effective at affecting product sales, and is insensitive to the discount factor.[3]

The rest of the paper is presented as follows. In Section 2, we discuss related work. In Section 3 we present the exclusion restrictions, and Section 4 develops our identification proof. We describe our empirical application in Section 5, while the counterfactuals are described in Section 6. Section 7 concludes.

2 Review of Literature

Proofs of identification for dynamic discrete choice models often build on the conditional choice probability approach introduced in Hotz and Miller (1993), who assume that all state variables that are observed to the researcher, and there is no unobserved heterogeneity across consumers.

[3]The higher effectiveness of promotional depth relative to frequency has been found in prior work by Osborne (2018b) under the assumption of a fixed discount factor.

In this setting, under a set of regularity conditions on the error term, one can flexibly estimate a consumer's *choice specific value*, which is the sum of the current period flow utility and the discount factor multiplied by the value function. The choice specific values are identified conditional on a normalization of the utility of one alternative (typically called the reference alternative), and given the functional form of the error distribution. With no restrictions on the functional form of the flow utility, the discount factor is not identified: In the conditional choice probability approach, one can roughly think of each moment as corresponding to the probability of a consumer choosing each alternative at each value of all the state variables. A fully flexible model would allow the utility function to be unique for each alternative and each state. Hence, if the discount factor were fixed, the number of moments and unknowns would be equal, and the model would be exactly identified. Formally, to identify the discount factor, some restrictions must be put on the utility function. Such restrictions will reduce the number of parameters in the model to be smaller than the number of moments, allowing the discount factor to be identified.

One type of restriction that has been proposed to help identify the discount factor is called an *exclusion restriction.* As explained in the introduction, this restriction requires the dynamic model to have at least two values of state variables, where for some choice alternatives, the current flow utilities remain unchanged but the expected future value could differ. Magnac and Thesmar (2002) is widely cited as the first paper which shows how exclusion restrictions can identify the discount factor. However, it should be pointed out that their exclusion restriction is defined in a way that is quite different from the definition that we use here. It is difficult to give an economic interpretation to the exclusion restriction used in Magnac and Thesmar (2002). Fang and Wang (2015) were the first to characterize the definition of the exclusion restriction in the way that we use. More recently, Abbring and Daljord (2018) show that the Fang and Wang (2015) exclusion restrictions may not allow for the discount factor to be point identified.[4] In our setting, we show that we can obtain point identification under the case of multiple exclusion restrictions (Abbring and Daljord (2018) also show this in the context of other models).

To our knowledge, in the empirical literature there are only a handful papers that explore such an identification argument to estimate consumer's discount factor or her incentive to consider future payoffs (Chung, Steenburgh, and Sudhir (2013), Lee (2013), Ching, Erdem, and Keane (2014), Chevalier and Goolsbee (2009), Ching and Ishihara (2018), Ishihara and Ching (2012)). These papers investigate sales force compensation schemes, rewards programs, consumer learning,

[4] Fang and Wang (2015) discuss identification of three discount factor parameters: the geometric discount factor, hyperbolic discounting, and naïveté. Abbring and Daljord (2018) show that the case of geometric discounting is a singular case in Fang and Wang (2015)'s proof, and present a proof of identification for this case.

and how the price of used goods affects the demand for new goods. As far as we know, none of the published research on structural models of consumer stockpiling have attempted to estimate the consumer discount factor (e.g., Erdem, Imai, and Keane (2003), Hendel and Nevo (2006a), Chan, Narasimhan, and Zhang (2008), Seiler (2013), Liu and Balachander (2014), Pires (2016), Sun (2005)). This is probably because previous structural models on consumer stockpiling all have assumed that the storage cost is an increasing and continuous function of inventory. This simplifying assumption, though convenient, has ruled out the exclusion restrictions that we use in our identification arguments. As a result, all previous structural empirical work on consumer stockpiling fixes the discount factor to be consistent with the interest rate, instead of estimating it.[5] Our paper is the first to argue that by properly modelling storage cost as a step function of inventory, this key state variable provides natural exclusion restrictions that can help identify the model's parameters, including the discount factor.

On-going research by Akça and Otter (2015) describes an alternative mechanism by which inventory can be used to identify the discount factor. They show theoretically that if consumers use brands within inventory in a last-in-last-out order, consumption rate are constant, and there is no unobserved heterogeneity across individuals, the discount factor can be identified. Although they do not estimate discount factors from field data, they present some survey evidence that could support the last-in-last-out assumption: If individuals are holding multiple packages of a product, they prefer to finish the one that is already open rather than opening another one. In contrast to the approach that we present, the approach of Akça and Otter (2015) has advantages and disadvantages. One advantage of our approach is that it is likely computationally easier to implement in practice. To implement the Akça and Otter (2015) approach, one would have to track an individual's inventory composition in terms of brands. In many product categories, there are many different brands available (for example, in our empirical exercise we include 19 brands), and it would be infeasible to track a state space with many brands. In contrast, if one implements our approach, at most one needs to track the composition of inventory in terms of package sizes held, which is much more tractable. In many packaged goods categories individuals primarily choose from 3 to 5 different package sizes. Another advantage of our approach is that it does not rely on any particular assumption related to the order of consumption of brands within inventory. In

[5] Note that with the assumption that the storage cost is an increasing and continuous function of inventory, a consumer has an incentive to wait longer before buying a new bottle, since the storage cost keeps dropping as inventory shrinks. This has the opposite effect of the increase in expected stock-out cost as the inventory drops. Therefore, the models used in previous work do not have clear implications about consumer purchase behavior as inventory drops.

particular, if one is willing to assume that storage costs are zero, one needs to make no assumption on the order of consumption. If storage costs are increasing, then it is necessary to make some assumption about the order in which different package sizes are consumed, but that assumption can be anything the researcher wishes - it is not limited to last-in-last-out. A potential advantage of Akça and Otter (2015)'s approach over ours is that it can be applied to product markets where storage costs do continuously increase as inventory increases, such as potato chips, rice or dried beans.

We note that in both our paper, and in Akça and Otter (2015), our theoretical proofs of identification rely on all states being observed to the researcher, and on the absence of unobserved heterogeneity across consumers, which is standard in the literature. Although in applications inventory is technically unobserved, since consumption is unobserved, if the assumption of constant consumption rates is reasonable then the consumption rate can be estimated separately from the model, and one can impute inventory given an assumption about initial inventory. As a result, one can treat inventory as observed. Nevertheless, to show that identification can still be achieved even if inventory cannot be imputed (for example, if consumption rates are stochastic), in Online Appendix L, we present evidence from numerical simulation and solution of the model with unobserved inventory that the shape of the purchase hazard can lead to identification of the discount factor. We also show in monte carlo exercises that one can still estimate the discount factor from simulated data. With respect to unobserved heterogeneity, one can technically identify individual-level heterogeneity if one observes a long panel of purchases, as we do (we estimate our model on 3 years of scanner data). To show that the heterogeneity can be identified in practice, we perform an artificial data experiment (Online Appendix G) where we simulate purchases given our estimated parameters from field data, and show our estimation procedure recovers the mean and standard deviations of the underlying parameters.

Finally, it is worth noting that Geweke and Keane (2000), Houser, Keane, and McCabe (2004) and Yao, Mela, Chiang, and Chen (2012) explore another identification strategy which requires the current payoffs are either observed or can be recovered from a static environment first. Yao, Mela, Chiang, and Chen (2012) assume consumers solve a dynamic programming problem and use this strategy to estimate the discount factor. Because Geweke and Keane (2000) and Houser, Keane, and McCabe (2004) do not assume consumers solve a dynamic programming problem, they recover the expected future payoffs but cannot separately identify the discount factor.

3 Exclusion Restrictions in the Stockpiling Model

In this section we develop a model of consumer stockpiling that is simplified somewhat from the model we will use for our empirical application, but contains its most important features. We use this model to show how our exclusion restrictions can be used to identify all of the model parameters.

We assume that the econometrician observes a market containing N consumers making purchase decisions over T periods. Consumers are forward-looking and discount the future at a discount rate $\beta_i < 1$. In this stylized model, we assume that a single product is available to consumers in some discrete package size. Each decision period t is broken up into two phases which happen sequentially: (i) a purchase phase, and (ii) a consumption phase.

In the purchase phase, consumer i observes her inventory (I_{it}), the price of a package of the product (p_{it}), an exogenous consumption need (c_i), and a choice-specific error (ε_{ijt}). The consumer's choice is her decision of how many packages of the product to buy, which we denote as $j \in \{0, 1, ..., J\}$. After making her purchase, the consumer enters her consumption phase, and receives her consumption utility.

We denote the size (or volume) of a package as b, and for simplicity of exposition we assume that b is an integer. We denote the consumer's inventory (which will also be integral) at the beginning of the period as I_{it}. We also make the assumption of integral inventory for convenience in the theoretical sections; it is straightforward to extend the model to continuous inventory, which we do in the empirical application. We assume that consumption rates are constant over time (but may vary across individuals), which is an assumption that we maintain in the empirical model. We note that this assumption makes our proofs of identification simpler, but is stronger than necessary (in particular, we discuss the case of unobserved and time-varying consumption rates in Online Appendix L). An advantage of the assumption of constant consumption rates is that they can be estimated from purchases prior to estimation of any empirical model; we provide support for this approach in our application in Online Appendix B. We assume that consumption rates are integral and lie in the set $\{0, 1, 2, ..., \overline{c}\}$. If the consumer's inventory at the end of the purchase phase, which we denote as $I_{it} + b \cdot j$, is above the consumption need c_i then she receives consumption utility γ_i. If she cannot cover her consumption need then she incurs a stockout cost ν_i.[6] At the end of the period, the consumer incurs a storage cost $s(\cdot; \boldsymbol{\omega}_i)$. Inventory will generate exclusion restrictions if

[6] We assume that the stockout cost does not depend on the consumption need but this assumption is innocuous. We could also assume that the stockout cost is proportional to the difference between inventory and the consumption shock, and our identification results will be unaffected.

there are some values of I for which $s(\cdot;\boldsymbol{\omega}_i)$ does not change. Our first assumption related to the exclusion restrictions, stated below, formalizes this restriction on storage cost.

First Model Assumption Related to Exclusion Restrictions, X1

1. The storage cost function s is only a function of the number of packages held at the end of the period, B, rather than inventory I, and the package size $b > 1$.

The number of packages held can be written as the following function of inventory: $B_{i,t+1}(j, I, c_i) = \lceil \max\{(I_{it} + b \cdot j - c_i)/b, 0\} \rceil$.[7] The assumption that $b > 1$ ensures that X1 is meaningful. $\boldsymbol{\omega}_i$ is a vector of parameters determining how storage costs vary with the number of packages held. For now, we allow the storage cost function to be non-parametric:

$$s(B;\boldsymbol{\omega}_i) = \omega_{i,B}. \tag{1}$$

We will assume that the cost of storing 0 packages is 0. In practice, one may consider imposing a functional form on s. One possibility is that the storage cost could be close to zero up to some limit, and then it goes up dramatically. For instance, many households' laundry room has reserved space for a few bottles of laundry detergent. In general, there will be a limit on how many packages a household can store, and we denote this limit as M.

The assumption that a consumer's storage cost depends on the number of packages held is valid for many product categories. For example, products that are sold in bottles or boxes such as laundry detergent or breakfast cereal will likely satisfy this assumption; products such as potato chips where the package size can be shrunk as the product is used up would likely not. The cost to storing laundry detergent depends on the amount of space taken up by the bottle, but not the amount of liquid within the bottle.

We note that Assumption X1 will help generate exclusion restrictions, but on its own it is not enough. Our exclusion restrictions would be violated if utility were continuously increasing in consumption, as individuals would optimally consume more at higher levels of inventory. To ensure that consumption utility does not increase with inventory, we make a second assumption as follows:

Second Model Assumption Related to Exclusion Restrictions, X2

2. The consumption need is not a function of inventory.

Intuitively, this assumption says consumers receive no additional utility from consuming more than their consumption needs. A limitation of our work is that this assumption can only be applied

[7]The ceiling function $\lceil \cdot \rceil$ returns the smallest integer that is greater than or equal to its argument.

to a limited number of product categories, in particular, those which an individual would use on a regular basis or as a need arises. Such product categories could include cleaning products such as dishwashing or laundry detergent, or condiments such as ketchup or mustard (it is unlikely that people derive additional utility from putting more mustard than necessary on a hot dog, for example). In particular, this assumption may not hold for product categories such as snacks or sweets where the temptation to consume may rise with larger inventory stocks (e.g., Chan, Narasimhan, and Zhang (2008), Sun (2005)).

Even if we maintain Assumptions X1 and X2, our model may not generate exclusion restrictions if the consumption need, c_i, only took on values that were multiples of the package size, as storage costs would always change when inventory changed. To guarantee the existence of exclusion restrictions, the researcher would need to observe consumer's choice for at least three values of inventory per package size where storage costs do not change.

Third Model Assumption Related to Exclusion Restrictions, X3

3. There are at least three inventory levels per package size where storage costs do not change, and the maximum number of packages an individual can hold is at least two: $\bar{c} \leq b/3, M \geq 2$.

The restriction in X3 will ensure that for most inventory levels consumers face, the number of packages held will remain unchanged (and hence storage costs will also stay the same). We note that in our application, it needs to be the case that the exclusion restriction holds for at least three levels of inventory in order to generate point identification of β, and that the researcher can observe choice probabilities when either one or two packages are held, which we will prove in Section 4 below. We note that if $\beta_i = \beta$ for all consumers, then the assumption that $\bar{c} \leq b/3$ is stronger than necessary for identification. In that case, it would simply need to be the case that there is one level of the consumption rate, c_i, where $c_i \leq b/3$ and a sufficient number of individuals have a consumption rate c_i that choice probabilities can be estimated for all individuals at these inventory levels. Exclusion restriction X3 is likely to hold for product categories such as laundry detergent, ketchup, etc., where individual consumption needs are a small fraction of a package size. It would likely be violated for product categories where every time consumption occurs, a package is used up (such as canned tuna or canned soup).

Given this information, we can write down the consumer's flow utility as follows:

$$\begin{aligned} & u_{it}(j, I_{it}, \varepsilon_{ijt}, p_{it}, c_i; \boldsymbol{\theta}_i) \\ &= \begin{cases} \gamma_i - s(B_{i,t+1}(j, I, c_{it}); \boldsymbol{\omega}_i) - \alpha_i p_{it} j + \varepsilon_{ijt} & \text{if } I_{it} + b \cdot j \geq c_{it} \\ -\nu_i - \alpha_i p_{it} j + \varepsilon_{ijt} & \text{otherwise} \end{cases}, \end{aligned} \tag{2}$$

where $\boldsymbol{\theta}_i = (\alpha_i, \beta_i, \gamma_i, \nu_i, \boldsymbol{\omega}_i)$ is a vector of the consumer utility coefficients and the discount factor, ε_{ijt} is a choice-specific error, α_i is the price coefficient. Additionally, γ_i and ν_i cannot be separately identified. Hence, we normalize $\gamma_i = 0$.[8]

We assume that consumers believe that the product's price follows a stochastic Markov process with a transition density $F(p_{i,t+1}|p_{it})$. Denoting the vector of choice-specific errors as $\boldsymbol{\varepsilon}_{it}$, the consumer's Bellman equation can be written as follows:

$$V_{it}(I_{it}, p_{it}) = E_{\boldsymbol{\varepsilon}_{it}} \max_{j=0,\ldots,J} \{u_{it}(j, I_{it}, \varepsilon_{ijt}, p_{it}, l; \boldsymbol{\theta}_i) + \beta_i E_{p_{i,t+1}|p_{it}} V(I_{i,t+1}, p_{i,t+1})\}. \tag{3}$$

The transition process for the inventory state variable I_{it} is

$$I_{i,t+1} = \max\{I_{i,t} + b \cdot j - c_i, 0\}.$$

We assume that if a consumer makes a purchase when her inventory is above $Mb - c_i$, then her inventory is set to the upper bound Mb. Intuitively, this is consistent with a situation where a consumer's storage space is used up, but if she purchases another bottle, she takes the one that is already open and gives it away or otherwise disposes of it.

4 Theoretical Identification: Proof

In this section, we present a constructive proof of the identification of the model presented in Section 3. Our proof relies on the researchers being able to estimate choice probabilities as a function of states, and shows that one can derive formulas for all the model parameters in terms of these choice probabilities. We make six additional assumptions below:

Assumptions A1-A6

1. In a given purchase occasion, the maximum number of packages a consumer is allowed to buy is 1.

2. Prices are fixed over time at a level p.

3. All model parameters are the same for all individuals, and the consumption rate, $c_i = 1$.

4. The choice-specific error term, ε_{ijt}, follows a type-1 extreme value distribution.

5. Inventory, I_{it}, is observed to the researcher.

[8] It can be shown that if we reduce γ_i by $\epsilon > 0$ and add ϵ to ν_i, the difference in choice-specific utility remains unchanged at all inventory states.

6. The stockout cost, ν, is positive.

We make A1 and A2 for convenience. It would be straightforward to extend our proof to allow individuals to choose more than 1 package, and to incorporate price transitions, but the notation would be more complicated. Assumptions A3-A5 are standard assumptions used in the literature on identification of dynamic discrete choice models (Fang and Wang (2015), Abbring and Daljord (2018) and Akça and Otter (2015) all maintain these assumptions). Unobserved heterogeneity can be included if the researcher has a panel dataset where the time dimension is large relative to the panel dimension. In that case, a consistent estimate of individual-level choice probabilities could be obtained, and one could compute individual-level parameter estimates. Assumption A4 allows us to obtain closed-form solutions for the model parameters in terms of choice probabilities, but can be relaxed as long as the error distribution is well-behaved (Rust (1994), Magnac and Thesmar (2002)). A5 is a standard assumption, which is reasonable if consumption rates are constant and some estimate of inventories at the beginning of the researcher's estimation sample can be constructed. A6 is necessary for stockpiling to occur - if $\nu = 0$, individuals will not stockpile and purchase probabilities will always be constant. In Online Appendix L, we provide numerical evidence on how identification may still be obtained if A5 is relaxed (i.e., consumption rates are stochastic). We note that our discussion in Online Appendix L is not a theoretical proof of identification with unobserved states; such a proof is beyond the scope of this paper.

Below, we define a rank condition that is necessary for identification of β.

Rank Condition R1

Let $\hat{P}(I)$ denote the empirical probability of purchase at inventory level I. There exists two levels of inventory, I and $I+1$ where the exclusion restrictions hold and $\hat{P}(I) \neq \hat{P}(I+1)$, and

$$\begin{aligned}
&\log((\hat{P}(I+2)) - \log(\hat{P}(I+1)) - (\log(\hat{P}(I+1)) - \log(\hat{P}(I))) + \\
&\log(1+(1-\hat{P}(I+b+1))/\hat{P}(I+b+1)) - \log(1+(1-\hat{P}(I+1))/\hat{P}(I+1)) - \\
&\log(1+(1-\hat{P}(I+b))/\hat{P}(I+b)) - \log(1+(1-\hat{P}(I))/\hat{P}(I)) \neq 0.
\end{aligned} \tag{4}$$

Theorem 1 *Suppose Assumptions A1-A5 and our exclusion restrictions X1-X3 hold. If rank condition R1 holds, then the parameters of the stockpiling model, α, β, ν, and ω_1 through ω_M, can be globally identified.*

Proof. Denote $\hat{P}(I)$ as the observed probability of purchase at inventory level I. Define v_j to be the choice-specific value of buying j packages at inventory level I and parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \nu, \omega_1, ..., \omega_M)$:

$$v_j(I;\boldsymbol{\theta}) = -\alpha p\mathbf{1}\{j=1\} - \omega_{B(j,I,1)} - \nu\mathbf{1}\{I=0\} + \beta V(\max\{I+bj-1,0\}). \tag{5}$$

Under the logit error assumption, we can write the choice probabilities in terms of choice-specific values as follows,

$$\Delta\log(\hat{P}(I)) \equiv \log(\hat{P}(I)) - \log(1-\hat{P}(I)) = v_1(I;\boldsymbol{\theta}) - v_0(I;\boldsymbol{\theta}). \tag{6}$$

If a consumer can hold up to M packages, then the number of parameters we need to identify is $M+3$: these are the M different values of ω_B, the stockout cost ν, the discount factor β, and the price coefficient α. Below, we show that all the parameters of the model can be expressed in terms of choice probabilities. Details of the derivations are provided in Online Appendix A.

Suppose first that Rank condition R1 holds. We show in Online Appendix A that rank condition R1 implies that $\beta > 0$.[9] If we define

$$\hat{\Phi}(I) \equiv \log\left(1 + \frac{\hat{P}(I+b)}{1-\hat{P}(I+b)}\right) - \log\left(1 + \frac{\hat{P}(I)}{1-\hat{P}(I)}\right),$$

then for $\beta > 0$, we can express β as

$$\hat{\beta} = \frac{\Delta\log(\hat{P}(I+2)) - \Delta\log(\hat{P}(I+1))}{\Delta\log(\hat{P}(I+1)) - \Delta\log(\hat{P}(I)) + \hat{\Phi}(I+1) - \hat{\Phi}(I)} \tag{7}$$

for a value of $I > 1$ and I such that storage costs do not change over the interval I through $I+2$. The exclusion restrictions assumptions (X1-X3) will guarantee that such an interval can be found, and rank condition R1 implies that a solution for β exists since Equation (4) is not 0. The price coefficient, α, can be derived from the purchase probability when an individual's inventory reaches the capacity constraint:

$$\hat{\alpha} = -\frac{\Delta\log(\hat{P}(Mb))}{p}.$$[10]

Given ($\hat{\beta}$, $\hat{\alpha}$), the stockout cost, ν, and the storage cost for one package, ω_1, can be expressed as

[9] The case of $\beta = 0$ is straightforward. If choice probabilities are flat for all inventory levels belonging to the same package, we can infer that $\beta = 0$.

[10] We note that the identification of α arises from the assumption that when an individual has M packages in inventory she disposes of the package that is currently begin used and sets her inventory level to Mb. In general it would be preferable to obtain identification of the price coefficient from price variation, rather than an assumption about how inventory is filled up when a consumer reaches her maximum storage capacity.

the solution to the system of linear equations,

$$\begin{bmatrix} \frac{\hat{\beta}^{b-1}-1}{1-\hat{\beta}} & \frac{2\hat{\beta}-\hat{\beta}^{b}-1}{1-\hat{\beta}} \\ \frac{\hat{\beta}^{b}-1}{1-\hat{\beta}} & \frac{\hat{\beta}-\hat{\beta}^{b+1}}{1-\hat{\beta}} \end{bmatrix} \begin{bmatrix} \hat{\omega}_1 \\ \hat{\nu} \end{bmatrix} = \begin{bmatrix} \hat{\alpha}p - \Delta\log(\hat{P}(0)) + \hat{\beta}h_0(\hat{\beta}) \\ \hat{\alpha}p - \Delta\log(\hat{P}(1)) + \hat{\beta}h_1(\hat{\beta}) \end{bmatrix}, \tag{8}$$

where the terms h_0 and h_1 are functions of β and choice probabilities $\hat{P}$ for inventory values between 0 and b (these functions are defined in Online Appendix A). Note that the above system of equations is derived from the purchase probabilities at inventory levels of 0 and 1. Higher values of storage costs such as ω_2 through ω_M can be derived from the choice probabilities at inventory levels 2, $b+2$, $2b+2$, and so on. In particular, one can derive $\hat{\omega}_2$ as

$$\hat{\omega}_2 = -\hat{\alpha}p + \hat{\omega}_1 + \hat{\beta}(V(b+1) - V(1)) - \Delta\log(\hat{P}(2)), \tag{9}$$

where we show in Online Appendix A that one can express $V(b+1)$ and $V(1)$ in terms of choice probabilities, and the parameters $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\omega}_1$. In general,

$$\hat{\omega}_B = -\hat{\alpha}p + \hat{\omega}_1 + \hat{\beta}(V((B-1)b+1) - V((B-2)b+1)) - \Delta\log(\hat{P}((B-1)b+2)), \tag{10}$$

where the value function difference $V((B-1)b+1) - V((B-2)b+1)$ is a function of choice probabilities and parameters that we have already solved for α, β, ν, ω_1, ..., ω_{B-1}.

■

4.1 Theoretical Identification: Discussion

In this section, we provide some additional comments on the interpretation of equation (7) in Theorem 1, which defines β, and discuss the implications of relaxing the assumptions of constant consumption rates and observed inventory. First, we note that although global identification potentially fails if rank condition R1 fails, this occurence should be a knife-edge case that would occur with probability zero in actual datasets. We provide support for rank condition R1 in our empirical application in Section 5.3.

We provide details on how the formula for β in equation (7) is derived in Online Appendix A. Importantly, in the derivation we show that in order to express β only in terms of choice probabilities, we need to solve two moment conditions,

$$\begin{aligned} \Delta\log(\hat{P}(I+1)) - \Delta\log(\hat{P}(I)) &= \beta(V(I+b) - V(I) - (V(I+b-1) - V(I-1))) \quad (11) \\ \Delta\log(\hat{P}(I+2)) - \Delta\log(\hat{P}(I+1)) &= \beta(V(I+b+1) - V(I+1) - (V(I+b) - V(I))), \quad (12) \end{aligned}$$

where the exclusion restrictions hold for $I, I+1$, and $I+2$, and the value function differences can be expressed in terms of β and choice probabilities. In the standard formulation of the exclusion restrictions, one would invert a single one of the above equations (i.e., equation (11) at a particular value of I) to arrive at β. Abbring and Daljord (2018) have recently shown that for general discrete choice dynamic problems, multiple values of β can be a solution to a single moment condition (resulting from a single exclusion restriction), an issue that arises in our setting as well. To show this, in the left panel of Figure 1 we plot an individual's expected future value of making a purchase, $\beta(V(I+b)-V(I))$, as a function of inventory at different β values, for a numerical solution of the model in Section 3.[11] Note that for a given value of β, the slope of the expected future value curve corresponds to the right hand side of equation (11).[12] If one were to use a single moment condition to solve for β, one would compute the value of the left hand side of equation (11) at some level of inventory, and then find a β where the slope of the expected future value curve is equal to that value. However, it can be seen from the figure that this slope is not monotonic in β: for low levels of β, the slope gets larger in magnitude as β rises, but for high values of β it decreases in magnitude. This non-monotonicity suggests that two values of β can solve a single moment condition. In Online Appendix A, we show that by using two moment conditions (differencing equations (11) and (12)), we can rule out one of the solutions for β.

The formula for β derived in equation (7) has an intuitive interpretation: It suggests that β is an increasing function of the ratio of how much purchase probabilities decrease with inventory at high levels of inventory, as compared to low levels of inventory. To support this interpretation, in the right panel of Figure 1, we plot the theoretical purchase probabilities as a function of inventory and β for the same model parameterization. The purchase probability curve flattens out as inventory increases because the disutility of a future stockout is discounted more for higher inventory values. For a relatively myopic individual, this disutility will not become apparent until inventory levels are low, and will be significantly discounted at high levels of inventory. As a result, the purchase probability curve flattens out more quickly for a more myopic individual. For a forward-looking individual, the disutility from a future stockout will matter even at higher levels of inventory, resulting in less curvature of the purchase probability function. Lower curvature of the purchase

[11] We set $p = 3.31$, $\nu = 0.33$, $\alpha = 1$, $b = 8$, $M = 3$, and storage costs to zero, corresponding to the median values of the parameter estimates from our empirical exercise. The value of p corresponds to the average price of the most popular size. The value $b = 8$ is in line with the amount of time it takes an average consumer to use up the most popular size of detergent in the data. See Online Appendix B for more details.

[12] In Online Appendix K, Propositions 1 to 3, we provide proofs that the expected future value of a purchase decreases with inventory, and increases with β when storage costs are sufficiently small.

probability function will bring the numerators and denominators of equation (7) closer together, producing an estimate of β that is closer to 1. The patterns we show in Figure 1 are robust to increasing the size of storage costs (see Online Appendix Figures 5 and 6), increasing the weight on the error term relative to utility (Online Appendix Figure 7), and allowing for individuals to purchase difference package sizes (Online Appendix Figure 9).

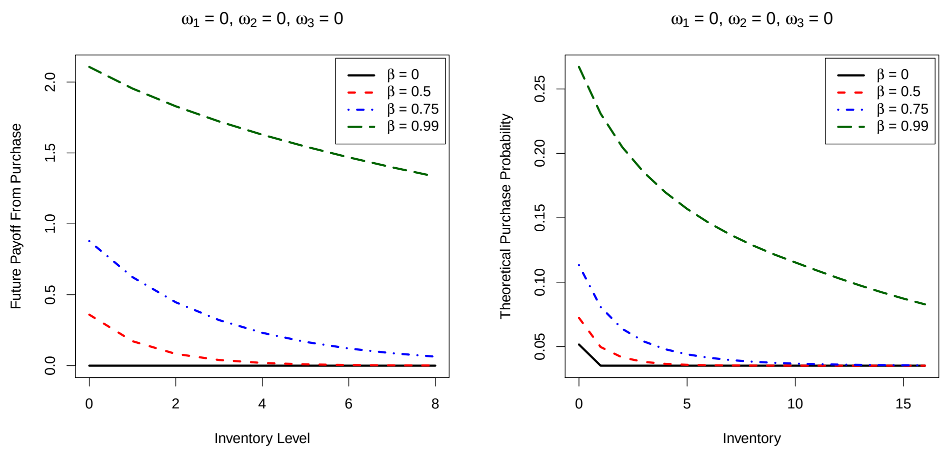

Figure 1: Expected future payoff from purchase, $\beta\,[V(I_1'(I)) - V(I_0'(I))]$ (left panel), and theoretical purchase probability (right panel), as a function of I and β, for storage costs of zero. Parameter values $\nu = 0.4$, $M = 3$, $p = 3.31$, and logit error term. Note that $I_1'(I) \equiv I - 1 + b$ is an individual's next period inventory if a purchase is made, and $I_0'(I) \equiv \max\{I-1, 0\}$ is future inventory if no purchase occurs.

Before proceeding to the empirical application, we comment briefly on identification in situations where inventory is unobserved to the researcher. In our empirical application, we assume that consumption rates can be estimated outside of the model, and given we assume that inventories are zero several years prior to the beginning of the estimation data, we can construct an estimate of a household's inventory level every week. However, in applications where consumption rates are stochastic, such an imputation may not be possible, meaning that it will not be possible to estimate choice probabilities at different inventory states. In Online Appendix L, we investigate the identification of a similar version of the stockpiling model with stochastic consumption rates and unobserved inventory using numerical simulations.

To summarize, we find that the purchase hazard, which is the average probability that an individual makes a purchase, given the last purchase occurred τ periods in the past, has similar properties to the purchase probability graph in the right panel of Figure 1. Intuitively, time since last purchase is a proxy for inventory, and the purchase hazard will flatten out as β increases.

Our numerical results also suggest that there may be a potential identification failure between consumption rates and the stockout costs, as these two parameters have similar effects on the purchase hazard when β is small. This result provides an additional justification for our assumption that we should calibrate consumption rates outside the model. We also provide a series of artificial data experiments (Online Appendix L.3), where we simulate a dataset with unobserved inventory and stochastic consumption rates, and show that we can recover the underlying model parameters. We note that the exclusion restrictions are still likely to be useful in helping with identification even if there are some unobserved states, as they restrict the number of parameters available to fit the moments that are observed to the researcher (the purchase hazard). A formal identification proof is beyond the scope of this paper and we leave it to future research.

5 Empirical Application

5.1 Data and Sample Construction

To quantify the substantive implications of identifying the discount factor, we estimate a stockpiling model using individual level IRI data in the laundry detergent category (Bronnenberg, Kruger, and Mela 2008). An observation in our data is a household-week pair. We observed individual purchases between the years 2001 through 2007. The final 3 years of the data are used to estimate the model parameters (our model likelihood is constructed for these years), while the first 4 are used to construct initial inventories. In our sample we include households who only purchase the 5 most popular sizes of detergent: the 50 oz, 80 oz size, 100 oz size, the 128 oz size, and the 200 oz size. We restrict the sample to include households who purchase from the top 25 brands by overall purchase share. We also allow consumers to purchase up to 5 bottles of a size, because sometimes individuals will purchase multiple bottles of a given brand in a single week. We remove households who ever purchase different products within the same week (this is very infrequent), or who purchase more than 5 bottles of a product in a week. We also restrict the sample to include individuals whose purchase behavior looks like it is not missing purchases of detergent: a panelist might make a purchase of detergent, but not scan their receipt, meaning it would not be recorded in the data. To do this, we only include households who make at least 5 purchases between 2005 and 2007, and for whom the maximum number of weeks between purchases is smaller than 40 weeks. After applying these exclusions, we compute a household level consumption rate by computing the sum of total quantity over the seven year window, and dividing by the total number of weeks over which purchases are observed (some households enter or exit the data after 2001, or before 2007, so the length of the observed purchase history may not always be 7 years). We then impute inventory in

Table 1: Summary Statistics of Purchase Data

Number of households	312
Avg interpurchase time (weeks)	9.9
Fraction of weeks with 0 bottles bought	0.898
Fraction of weeks with 1 bottles bought	0.077
Fraction of weeks with 2 bottles bought	0.018
Fraction of weeks with 3+ bottles bought	0.025
Fraction of purchases where 100 oz size chosen	0.712
Fraction of purchases where 128 oz size chosen	0.136
Fraction of purchases where 200 oz size chosen	0.088
Fraction of purchases where 50 oz size chosen	0.037
Fraction of purchases where 80 oz size chosen	0.027

the period of the estimation sample by assuming inventory is zero in 2001, and calculating inventory forward using the imputed consumption rates. We exclude any households for whom we predict the household would stock out (have zero inventory) for 10 or more weeks in a row. After these exclusions, the estimation sample contains 312 households.

Some statistics on household purchase behavior are shown in Table 1. An average household makes a purchase about every 10 weeks, and in most weeks no purchase occurs. In our sample, the most frequently purchased bottle size is the 100 oz bottle, followed by the 128 oz bottle. Table 2 shows the purchase shares (the number bottles purchased of a particular brand divided by the total number of bottles purchased in the sample) as well as average prices (in cents per ounce) for each brand purchased by households in the sample (When constructing the sample we initially include the top 25 brands by purchase share. After reducing the sample to 312 households by removing those who purchase too infrequently or who purchase too much, only 19 brands have positive purchases). Prices vary significantly across brands, but also within each brand-size combination. We show that prices vary over time for a particular product (a brand-size combination) in two ways: by documenting price variation at the household level, and at the store level. At the household level, we compute the coefficient of variation of prices for all available brand-size combinations. The average within-household coefficients of variation are shown in Table 3. For most products, the standard deviation of prices that are observed by the household are about 15 to 20 percent of the average price, implying that households observe a substantial amount of price variation for each product from trip to trip.

Table 2: Brand Level Purchase Shares and Prices

Brand	Purchase Share	Price (Cents Per Ounce)
Tide	23.8	8.61
Xtra	8.9	2.48
Purex	9.2	4.81
All	7.9	5.64
Arm & Hammer	8.8	4.71
Era	5.9	5.25
Dynamo	11.7	4.58
Wisk	10.3	6.31
Private Label	3.8	3.51
Cheer	2.1	7.12
Fab	1.4	6.08
Yes	2.1	4.51
Ajax Fresh	0.4	3.19
Gain	0.5	6.16
Ajax	0.4	3.14
Trend	0.4	2.22
Sun	0.8	4.33
Solo	1.3	3.88
Ivory Snow	0.3	10.41

Notes: Purchase shares show the number of packages purchase of each brand in our trip-level data of 36,101 trips. The prices are the averages of prices, measured in cents per ounce, observed by a given household in a given trip, for brands that were available on that trip.

Similar amounts of variation in the data occur at the level of a store as well. Within a given store, a product's price tends to follow a pattern that is typical of many packaged goods in scanner data, where the price has a high regular price and periodically falls to a lower deal price for a short time period (for example, Online Appendix Figure 4 shows the time series path of the price of the 200 OZ size of Tide at a frequently-visited store). To quantify how often products go on deal at all stores in the data, for each available product we computed the average regular price, the average deal price, and the fraction of store-week observations where the product is on deal. When a product is put on sale, the deal price is typically 20 to 30 percent below the regular price, and deals typically happen in 20 to 30 percent of weeks (Online Appendix Table 19). In summary,

Table 3: Within Household Coefficient of Variation of Prices for Available Brand-Sizes

Brand	100 oz	128 oz	200 oz	50 oz	80 oz
Tide	13.81	-	14.65	13.9	13.03
Xtra	-	10.91	7.51	-	-
Purex	17.71	15.4	11.06	10.43	-
All	12.8	-	10.64	10.92	16.58
Arm & Hammer	16.94	-	11.45	-	-
Era	17.1	-	4.8	2.46	-
Dynamo	21.1	-	14.83	-	-
Wisk	14.3	-	15.86	-	17.42
Private Label	13.89	8.76	-	5.26	-
Cheer	-	-	-	-	7.76
Fab	11.73	-	-	18.82	-
Yes	19.14	-	-	-	-
Ajax Fresh	-	12.3	-	-	-
Gain	14.06	-	7.24	-	-
Ajax	-	11.8	-	-	-
Trend	-	2.14	-	-	-
Sun	24.52	-	-	-	-
Solo	21.79	-	-	-	-
Ivory Snow	-	-	-	6.71	-

Notes: The numbers in this table are constructed using our estimation sample, which was created by combining IRI's purchase panel, household trip panel, and store price panel. For each brand-size that exists in the sample, we compute at the household level the coefficient of variation. The estimates are calculated using 36,101 trips in our estimation data.

households will observe a significant amount of price variation over time for all products. Such price variation will encourage households to stockpile if they are forward-looking.

5.2 Estimation Details

This section outlines the estimation procedure used to recover the model parameters. Since we wish to include unobserved heterogeneity in many of our model parameters, we use the modified Bayesian MCMC algorithm proposed by Imai, Jain, and Ching (2009) to estimate the model.[13] Although we find evidence in Section L.3 that stockpiling models can in principle be identified under relatively flexible specifications of the storage cost function and distribution of consumption shocks, in practice such flexibility can greatly increase the computational burden of estimation. We have found two problems with including time-varying consumption shocks in the model. First, as we outline in Online Appendix L, the distribution of consumption shocks is difficult to separately identify from the stockout cost, and so we expect that the parameters governing the consumption shock draws would need to be normalized. Indeed, although Hendel and Nevo (2006a) incorporate normally-distributed i.i.d. consumption shocks in their stockpiling model, they do not estimated the mean and variance of the consumption shock process. Second, when we investigated incorporating such shocks into the model we found that modeling stochastic consumption shocks would add to the computational burden of estimation since the shocks need to be integrated out while estimating the other model parameters. As we use MCMC to estimate our model, we would need to add another Gibbs step to our estimation routine where we draw the consumption shocks for every individual and every period in the data. Moreover, for standard distributions of consumption shocks (such as a normal distribution) the posterior density of the shocks given the data will not have a form that is easy to draw from, necessitating the use of a Metropolis-Hastings step. Adding this step to the algorithm substantially slows down convergence.[14]

As a result, we assume that consumption needs may vary across individuals, but are fixed over time. We also note that this assumption with respect to consumption needs means that we can impute a household's inventory, given an assumption about initial inventories. We assume that consumer inventory is 0 at the beginning of the pre-estimation period in 2001; given this assumption and a household's estimated consumption rate we can calculate the household's inventory in any

[13] Hendel and Nevo (2006a) propose a three step method for estimating stockpiling models that uses maximum likelihood, but their approach cannot allow for unobserved heterogeneity across individuals. Ching, Imai, Ishihara, and Jain (2012) provides a practitioner's guide to the IJC approach.

[14] Some earlier empirical work on stockpiling such as Hendel and Nevo (2006a) and Sun (2005) has included stochastic consumption shocks, but those papers did not model unobserved preference heterogeneity as we do.

period. In Online Appendix B, we verify both of these assumptions are reasonable by i) showing that the distribution of imputed consumption rates across households looks realistic, and is correlated with household observables in ways that make sense, ii) showing that the imputed inventory measures make sense: first, most households hold around 3 bottles of detergent, and decreases in household inventory are highly correlated with purchase. This correlation becomes especially strong for low levels of imputed inventory.

The other issue we introduced in the previous paragraph related to the specification of the storage cost function and the way inventory is modeled. Significant computational complications arise in situations where individuals can choose among more than a single size of bottle or brand. In our dataset individuals choose among $X = 5$ different package sizes and $K = 18$ brands. Adding multiple brands and package sizes to the empirical model will increase the size of the state space. This is because one would need to: (i) track inventory and price for each brand; (ii) track the number of bottles of each size held in inventory, and model the order in which different sizes of bottles are consumed. The inventory composition would matter to the consumer since her storage cost will decrease as she uses up a bottle. A consumer who has two small bottles in her inventory will lower her storage cost more quickly than someone who has two large bottles. An additional complication is that multiple package sizes would require us to model the order in which packages are consumed. For instance, if a consumer has a large bottle and a small bottle in her inventory, we would have to decide whether she would use the small bottle before the large one, or vice versa. Below we will first describe how we handle the issues arising from including different bottle sizes, and then from including multiple brands.

We deal with the computational issues arising from incorporating multiple bottle sizes in two ways. Our preferred method is to assume storage costs are zero, which is the model we present as our main specification. Under this restriction on storage costs, it is not necessary for an individual to track inventory consumption in the state space, and it is not necessary to make any assumption regarding the order in which bottles are consumed in inventory. In an earlier version of the paper, we estimated a similar specification in which we imposed a bound on the amount of inventory an individual could hold. We found that our estimates of the bound were very high, and so in later versions of the work we have removed the bound.[15]

As a robustness check, in Online Appendix J we present estimates from an extended model where we allow for increasing storage costs. A detailed description of the model parameterization is presented in Section D. Our estimated storage costs from this version of the model suggest that

[15] In practice we need to bound the state space, and so we assume that an individual can hold at most 4,800 ounces of detergent, which is about 50 regular sized bottles.

for the vast majority of the population, storage costs are very low. In order for the model with increasing storage costs to be computationally tractable, we had to make a number of simplifying assumptions. First, we have to limit the number of bottles an individual can hold. We set this bound to 6 bottles, since our imputed inventory measures suggest that most individuals do not hold more than this (Online Appendix B). Second, we assume that bottles are used according to a first-in-first-out rule. Third, even with these assumptions, the number of possible bottle composition states an individual can reach is extremely large if there are 5 bottle sizes. To keep the size of the state space manageable, we bin the 100, 80 and 128 ounce sizes together (we assume that all of these sizes contain 100 ounces of detergent), and we treat states that are rarely visited as inadmissable (if an individual makes a purchase that would lead to such a state, we assume that the bottle purchased is removed from inventory, and impose a utility penalty). Because the estimates of the extended model are consistent with the zero storage cost model, we present the latter in the main paper because it is simpler, and does not rely on as many simplifying assumptions.

To deal with issues arising from including brand differentiation, we follow Hendel and Nevo (2006a) and make two simplifying assumptions: (i) consumers only care about brand differentiation at the time of purchase, and (ii) a form of inclusive value sufficiency modified from what Hendel and Nevo (2006a) proposed (the modifications we use were introduced in Osborne (2018a)). Assumption (i) means that all utility from consuming a particular brand arises when a consumer makes a purchase, and at the time of consumption only the overall level of inventory matters.[16] This implies that the composition of the inventory (in terms of brands) does not matter, and it drastically reduces the size of the state space. One interpretation of this assumption, propose in Hendel and Nevo (2006a), is that consumption is drawn equally from all packages held. The main model specification presented in the paper is consistent with this assumption, because it does not impose any order on consumption from different packages within inventory. Moreover, such an assumption may be behaviorally realistic in the laundry detergent market, since customers may purchase multiple brands for different purposes (such as washing colors, whites, baby clothes, etc). We assume that the flow utility received from a particular brand scales linearly with the number of packages purchased: the flow utility from purchasing j packages of size x and brand k is equal to $\frac{j}{J}\xi_{ikx}$, where one of the ξ_{ikx} coefficients is normalized to zero, and $J = 5$ is the total number of bottles an individual is allowed to buy.[17] The assumption that brand utility scales with the number of packages purchased

[16]Formally, assumption (i) means that the consumption utility, γ_i, does not depend on the brand purchased (as we argued earlier, the parameter γ_i is not identified so we normalize $\gamma_i = 0$).

[17]We note that for the normalized brand, the mean flow utility will not scale with j, but the flow utility from the idiosyncratic error and price disutility will (see Online Appendix E.3). The assumption that utility scales linearly

relates the inclusive value sufficiency assumption (ii), and we will show below that it will help reduce the size of the model's state space. The consumer's flow utility function from buying $j > 0$ units of size x of brand k can be written down as:

$$u_{it}(k, x, j, I_{it}, \varepsilon_{ijt}, p_{it}, c_i; \boldsymbol{\theta}_i) \tag{13}$$
$$= \begin{cases} \frac{j}{J}\xi_{ik} - s(B_{i,t+1}(j, I_{it}, c_i); \boldsymbol{\omega}_i) - \alpha_i p_{ixkt} j + \varepsilon_{ijxkt} & \text{if } I_{it} + b(x)j \geq c_i \\ \frac{j}{J}\xi_{ik} - \nu_i \frac{c_{it} - (I_{it} + b(x)j)}{c_i} - \alpha_i p_{ixkt} j + \varepsilon_{ijxkt} & \text{otherwise} \end{cases},$$

where $b(x)$ is the number of ounces in a bottle of size x. Before we write down the Bellman equation, we need to clarify the elements of the consumer's choice set. Consumers can either purchase nothing ($j = 0$), or purchase $j > 0$ units of a single brand-size combination (k, x). Denoting the feasible set of (j, k, x) combinations as C, the consumer value function can be written as:

$$V_{it}(I_{it}, \boldsymbol{p}_{it}) = E_{\varepsilon_{it}} \max_{(j,k,x)\in C} \{u_{it}(k, x, j, I_{it}, \varepsilon_{ijkt}, \boldsymbol{p}_{it}, c_i; \boldsymbol{\theta}_i) + \beta_i E_{p_{i,t+1}|p_{it}} V(I_{i,t+1}, p_{i,t+1})\}, \tag{14}$$

where p_{it} is a vector of brand-size level prices. Our second assumption of inclusive value sufficiency (IVS) simplifies the state space by assuming that rather than tracking individual prices, consumers track the expected flow utility arising from each available package size, which are the inclusive values. The standard formulation of IVS used in Hendel and Nevo (2006a) relies on the assumption of logit errors, and under their formulation the number of inclusive values equals the number of available package sizes, X, multiplied by the number of packages an individual can purchase, J.

Osborne (2018a) shows that the number of inclusive values one needs to track can be further reduced to only the number of available package sizes, X, under the assumption that flow utility scales with the number of packages purchased, coupled with an assumption that the choice specific error can be written in the form of a nested logit. The nesting structure has the choice of bottle size and number of bottles (x and j) in the outer nest, while the inner nest is the choice of brand, k, for a given bottle size. The inclusive value parameter for the brand choice nest is set to be $\frac{j}{J}$; division by the number of packages an individual can choose, J, is necessary for the inclusive value parameter to be between 0 and 1, which ensures the density of the error is well-behaved (Cardell (1997)). The nested logit formulation of the error just described, along with the assumption that the flow utility scales with j, allows the j parameter to be factored out of the expected brand-specific utility, and as a result the inclusive values do not depend on j.

---

with quantity purchased may be strong in some cases, especially if individuals purchase many packages at once, as marginal utility from additional purchases could start to decrease. In our setting, we think this is likely less of an issue since most people only buy 1 or 2 packages at a time.

Denoting C_1 as the set of feasible (j, x) combinations and $C_2(x)$ as the set of brands which are available in size x, the two aforementioned assumptions entail that an individual's expected utility over brands for choosing j packages of size x can be written as,

$$\frac{j}{J}\Omega_{it}(x) = \frac{j}{J}\ln\left(\sum_{k\in C_2(x)}\exp\left(\xi_{ixk} - J\alpha_i p_{ixkt}\right)\right).$$

Details on the above derivation are shown in Online Appendix C. To summarize, our implementation of IVS assumes that consumers track $\Omega_{it}(x)$, rather than each individual price p_{ixkt}. As a result, the Bellman equation in equation (14) can be written as

$$V(I_{it}, \mathbf{\Omega}_{it}) = \ln\left(\sum_{(j,x)\in C_1}\exp\left\{\frac{j}{J}\Omega_{it}(x) - \nu_i\frac{c_i - (I_{it} + bj)}{c_{it}}\mathbf{1}\{I_{it} < c_i\} + \beta_i E_{\mathbf{\Omega}_{it}|\mathbf{\Omega}_{i,t-1}} V(I_{i,t+1}, \mathbf{\Omega}_{it})\right\}\right), \quad (15)$$

where $\mathbf{\Omega}_{it}$ is an X-dimensional vector of inclusive values for all package sizes.

In addition to the different specification used for storage cost, we make three additional minor changes to the model specification from the specification used for artificial data experiments. First, we incorporate a fixed cost of purchase, FC_i, which is the disutility a consumer receives from making a purchase. We found it necessary to include this parameter in order to properly fit the low frequency of purchase we observe in the data.

Second, rather than estimating consumption needs we calibrate them from the data. Consistent with what we note in Online Appendix L.1, we found it difficult to identify both the consumption rate and the stockout cost together, and this problem seemed especially pronounced when the discount factor was low. Thus, we set each individual's consumption rate to the total quantity purchased over the estimation period, divided by the total number of weeks where the individual is observed. To ensure our results are not materially affected by this assumption, we perform a robustness exercise where we increase every individual's consumption rate by 25% and re-estimate the model.[18] We find our parameter estimates are relatively insensitive to the consumption rate. The estimated parameters from this specification are presented in Online Appendix J, Table 24.

In the data, there are some weeks where some consumers do not visit any store. During such weeks, the individual's inventory evolves, but does not make any purchase ($x = 0$). To capture this, the third change we have made is that we assume there is an exogenous probability a consumer goes

[18] If an individual always purchases more of a product at the time she runs out, which we might expect with necessities such as laundry detergent, the calibrated consumption rate will equal the underlying consumption rate. We have found in simulations that if stockout costs are low enough that individuals sometimes wait a few periods after running out to make a purchase, the calibrated rate somewhate understates the actual consumption rate.

to the store, which we estimate prior to estimating the other model parameters. This probability is incorporated into consumers' expectations when they update their value functions in equation (15).[19]

For simplicity of our exposition below, we outline how the solution of the model works when it is assumed that consumers always visit a store. In terms of unobserved heterogeneity, we include as much richness as possible. In particular, most of our product shifters are heterogeneous, as well as the price coefficient, the cost of stocking out, and the discount factor.[20]

The basic steps of the algorithm are as follows:

1. Draw the population-varying parameters using Metropolis-Hastings,

2. draw the means and of population-varying parameters,

3. draw the variance of population-varying parameters,

4. draw the population-fixed parameters using Metropolis-Hastings, and

5. update the value function.

We describe how we implement steps 1 to 4 in Online Appendix E.1, and step 5 in Online Appendix E.2. Some other details related to the construction of the inclusive value transition process and setup of the MCMC chain are described in Online Appendices E.3 and E.4.

5.3 Identification Assumptions and the Data

In this section, we discuss evidence that the assumptions necessary for the proof of identification in Section 4 are satisfied in our data. First, we note that the exclusion restriction assumptions X1 (discontinuous storage costs) and X2 (exogenous consumption rates) must be imposed from theory. X1 is true for our setting, since liquid laundry detergents are sold in plastic bottles of fixed sizes. X2 is also behaviorally sensible: Bottles of laundry detergent come with guidelines on how much to use when washing a load, and so individuals are unlikely to gain or lose additional utility from deviating from these guidelines.

[19] Occasionally, a household will make multiple trips in a single week wihtout buying detergent. Since our analysis is done at the weekly level, we make an assumption about which store is the household's focal store. We tabulate for each household how often they visit each store, and assume the store in which the laundry detergent purchase decision was made was the store that was visited more often.

[20] In general, in a random coefficients model at least some coefficients must be normalized to guarantee identification (Ruud 1996).

Table 4: Dynamic Parameter Estimates

Parameter	1st Tertile	Median	Mean	2nd Tertile
Price Coefficient	-0.29	-0.24	-0.27	-0.21
	[-0.31, -0.28]	[-0.26, -0.23]	[-0.29, -0.26]	[-0.22, -0.2]
Stockout Cost	0.29	0.39	0.48	0.5
	[0.24, 0.36]	[0.31, 0.49]	[0.37, 0.66]	[0.4, 0.67]
Discount Factor	0.62	0.94	0.71	0.99
	[0.14, 0.89]	[0.81, 0.98]	[0.58, 0.82]	[0.98, 1]
Fixed Cost of Purchase	-	-	-1.83	-
			[-1.91, -1.77]	
Log-likelihood	-19585.37			
Deviance Information Criterion	40510.98			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. Only the mean is shown for parameters that are fixed across the population. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Square brackets show 95% confidence intervals.

We provide evidence in support of X3 (at least 3 periods are necessary to use up a package) in Online Appendix B, where we present summary statistics on the imputed consumption rates for the households in our data. In particular, we find that the majority of households do at most 4 or 5 loads per week. The most popular sizes of detergent contain 32 or 64 loads, and so it will take individuals around 8 weeks to use up a bottle. The maximum consumption rate in the data is about 13 loads per week, and even for this consumption rate it will take a household five weeks to use up a 200 ounce bottle of detergent.

Identification of the stockpiling model also requires that Rank condition R1 holds, which is supported by the data. In particular, in Online Appendix B we regress a dummy variable for whether an individual makes a purchase in a given week on a flexible function of the measure of imputed inventory used in the structural model. Recall that we calculate imputed inventory by assuming that inventory is zero in an individual's first purchase 3 years prior to the beginning of the data set, and then compute inventory iteratively by adding purchases in a period, and subtracting consumption on a period.[21] The regression estimates show that choice probabilities decrease with

[21] This imputation will be correct as long as during the 3 year pre-estimation period, each individual runs out of inventory at some point. The reason for this is that the imputed inventory will be smaller than or equal to actual inventory in the initial pre-estimation week, and so whenever actual inventory is zero, imputed inventory will also be

inventory in regions of the state space where the exclusion restrictions should hold. For instance, in Table 8, in the region of the state space where an individual has 0 to 50 ounces, the estimated purchase probabilities decrease in inventory, and we know the exclusion restrictions hold for levels of inventory in this range because there are no bottle sizes smaller than 50 ounces. The regression estimates also provide support for Assumption A6 (positive stockout cost), as choice probabilities increase drastically when inventory is zero.

To verify Rank Condition R1 (equation (4)), we do a back-of-the-envelope calculation. Suppose that an individual holds 40 ounces of a 100 ounce bottle in inventory, and has a weekly consumption rate of 4 loads ($c = 12.5$ ounces) per week, and is considering purchasing another 100 ounce bottle (so $b = 100$). Then one can calculate the predicted purchase probabilities implied by the regression model in Table 8 at the inventory levels $I = 40$, $I + c = 52.5$, $I + 2c = 65$, $I + b = 140$ and $I+b+c = 152.5$ using the regression estimates, and an estimate of equation (4) in R1 is -0.07504.[22] Note that equation (4) is the denominator of the formula for β in equation (7). We can also use the implied choice probabilities from the regression to solve for a back-of-the envelope estimate of β from equation (7). The estimated probabilities at the aforementioned inventory values imply a β of 0.32. This is lower than the estimate of β derived from the structural model (the model controls for price variation and allows for purchases of different brands and bottle sizes), but it does suggest that individuals are more myopic than the rational expectations benchmark. We provide additional evidence that the empirical model can recover the model parameters, including the unobserved heterogeneity, in Online Appendix G, where we simulate purchases from the estimated model parameters, and then estimate the model on the simulated data and recover the underlying parameters.

5.4 Estimation Results

This section presents our estimation results. We run the MCMC sampler for 20,000 iterations, and drop the first 10,000 to reduce dependence on initial starting points. The MCMC algorithm appears to converge after about 2,000 to 3,000 draws, so our cutoff point is conservative (see Figure 15 of Online Appendix M). Table 4 shows the estimates of the main parameters that affect an individual's dynamic decision: the price coefficient, stockout cost, discount factor, and fixed cost of purchase. All of these parameters vary across the population, and are transformations of draws from a normal prior distribution with diagonal variance (the transformation applied to each parameter

zero.

[22] The predicted purchase probabilities at these values of inventory are $P(I) = 0.164$, $P(I+c) = 0.139$, $P(I+2c) = 0.132$, $P(I+b) = 0.11$, and $P(I+b+c) = 0.106$.

is discussed in Online Appendix E). We show the 33rd, 50th, and 66th percentiles as well as the population mean. To compute an estimated moment, for example the 33rd percentile of the price coefficient, first for each Gibbs draw we compute the 33rd percentile of the population distribution of parameter draws for the price coefficient. The estimated 33rd percentile is the average of the 33rd percentiles over all 10,000 saved Gibbs draws. The second row shows the 95% confidence bounds on each of the estimated moments. There is a significant amount of heterogeneity in all of these model parameters. Additionally, the magnitudes of the parameters indicate that price sensitivities, stockout costs, and fixed cost of purchase have significant effects on purchase behavior. A similar table showing the estimated parameter distributions for product-specific parameters are shown in Online Appendix Tables 20 and 21. We note that we include unobserved heterogeneity in most product-level taste coefficients. For some of the smaller share brands, we found that we had to make the coefficients homogeneous across the population. Additionally, for most products we separately estimate brand and size taste coefficients: an individual's taste for a particular product is the sum of the brand coefficient plus the size coefficient. We did this because we do not observe a lot of purchases for many products and so it would be difficult to separately identify such coefficients. For some of the larger share brands, such as Tide, we do include taste coefficients that are brand-size specific, as we can identify those coefficients and allowing additional flexibility improves the model's ability to fit purchase shares.

Turning to the discount factor, the population average of the weekly discount factor is about 0.71, which is much lower than the value of $1/(1+0.05)^{(1/52)} \approx 0.9995$ that one would calibrate from the annual interest rate of 5%. There is also a significant amount of heterogeneity in discount factors. The upper tertile of the distribution of discount factors are around 0.99, and close to the rational expectations benchmark. On the other hand, the lower tertile of the distribution of discount factors is 0.62, indicating a substantial number of individuals are myopic. This heterogeneity can also be seen in Figure 2, where we plot a kernel density of the average estimated discount factor for the population (for each individual, we compute the average of the discount factor estimate for all saved draws). The individual estimates suggest there is a mass of individuals who are forward-looking, and the rest of the population is spread out until the discount factor is about 0.1 or 0.05. Although our estimated discount factors are less than the rational expectations benchmark assumed in past work, low estimated discount factors are consistent with some other field studies that allow the parameter to be free (for example, Yao, Mela, Chiang, and Chen (2012) estimate in data on cellular phone usage that consumer discount factors are around 0.91). In the specification presented in the main text, we did not interact the heterogeneous parameters with demographic coefficients, because in artificial data experiments we found it difficult to precisely recover the coefficients of the demographic interactions. We present an auxiliary specification in Online Appendix J, Table 11 where we include demographic interactions. Most of the interactions are not statistically different from zero (expand on this when they are put in).

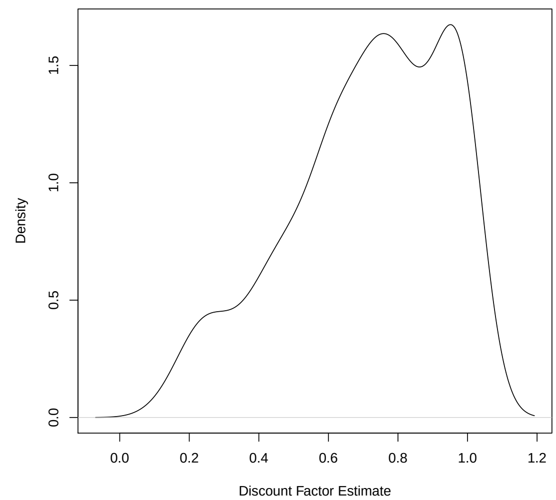

Figure 2: Kernel Density of Individual-Specific Discount Factor Estimates.

Taking our results at face value, it may be tempting to argue that our estimates suggest consumers are irrational, as a weekly discount factor of 0.71 would imply that consumers are close to being myopic when making financial decisions where the time horizon is on the order of a year. However, a discount factor in the range of our estimated value is also consistent with a setting where consumers think several weeks ahead when they make their purchase decisions.[23] Since consumer packaged goods are small ticket items, such a short planning horizon may be reasonable and could be rational behavior taking into account scarce mental resources. When making important financial decisions, consumers may behave in a more forward-looking way due to the fact that more money is at stake - there are larger gains from planning over a longer horizon and hence it is worth using more mental resources to think further ahead.

To understand how our results would change if we assume individuals are rational, we estimated the structural model again using a discount factor of 0.9995, the rational expectations benchmark. The estimates of the dynamic parameters are presented in Table 5. It is notable that the rational expectations benchmark produces a worse fit to the data, which can be seen in that is has both a lower log-likelihood and higher deviance information criterion. We provide additional comparisons of model fit in Online Appendix F, where we show that the main specification provides an improved

[23]It seems implausible that consumers plan years ahead for laundry detergent purchases.

Table 5: Dynamic Parameter Estimates: Discount Factor Fixed at 0.9995

Parameter	1st Tertile	Median	Mean	2nd Tertile
Price Coefficient	-0.31	-0.27	-0.29	-0.23
	[-0.33, -0.3]	[-0.28, -0.26]	[-0.3, -0.28]	[-0.24, -0.22]
Stockout Cost	0.27	0.33	0.36	0.39
	[0.24, 0.3]	[0.29, 0.37]	[0.32, 0.41]	[0.34, 0.45]
Fixed Cost of Purchase	-	-	-1.89	-
			[-1.98, -1.81]	
Log-likelihood	-19768.02			
Deviance Information Criterion	40773.41			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. Only the mean is shown for parameters that are fixed across the population. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Square brackets show 95% confidence intervals.

fit to interpurchase times, purchase probabilities given inventory levels, and price sensitivity. Interestingly, the other parameter estimates are similar, except that using the rational expectations assumption produces a slightly bigger stockout cost, which is intuitive. If individuals discount the future at a higher rate, the stockout cost will have a bigger effect on purchase likelihoods. As we will show in Section 6, the counterfactual predictions of the model can differ significantly as well if an incorrect discount factor is used. The driving force behind this difference is more forward-looking individuals will try harder to time their purchases to coincide with temporary price discounts.

In addition to the specifications discussed above, as additional robustness checks we also estimate a myopic version of the model, where the discount factor is set to zero, and a version of the model where we increase the calibrated consumption rates by 25%. The myopic model also provides a worse fit to the data than the model where we estimate the discount factor, with a marginal log-likelihood of -19629.16 and a deviance information criterion of 40560.51. The estimation results for the specification with higher consumption rates are presented in Online Appendix J, Table 24, with qualitatively similar results. The distribution of discount factors still suggests a lot of spread, with the upper quartile around 1. The average discount factor is a little lower, at 0.6.

Finally, in Online Appendix J, Table 11 we present estimates from a version of the model with increasing storage costs. As we describe in Online Appendix D, in order to make estimation of this version of the model tractable we must make a number of additional assumptions. In particular, we assume consumers can hold up to six bottles, that consumption order occurs according to a first-in-first-out rule, that states which are rarely visited in the data are inadmissable, and we bin together the 80 ounce and 128 ounce sizes together with the 100 ounce size. We bin the 80, 100 and 128 ounce sizes together to keep the states space of the model tractable. Recall that for each individual in the data, we track value functions on a grid of 100 price points; as we discuss in Online Appendix D, the number of inventory points we track is 17, which is smaller than the number used for the model without increasing storage costs. When we include an additional state to track the composition of inventories in bottle sizes, the size of the state space increases exponentially. In particular, if we were to include all 5 sizes, the number of states would be multiplied by 19,531. With only 3 bottle sizes, we only add 1,093 states. As we noted, we only track states that are visited in the data more than 5% of the time - this restriction reduces the number of states to 23, since many individuals repeatedly purchase only one or two different sizes. Even with our restrictions, we still must track 39,100 state space points for each household in the data, making this version of the model very computationally burdensome to estimate.

Additionally, to keep the number of parameters tractable, we assume that storage costs increase quadratically in volume held.[24] We emphasize that we do impose the exclusion restriction in the storage cost function in this model specification: storage costs are quadratic, but only change if a bottle is used up. The estimated parameters suggest that storage costs are very close to zero for most of the population. The storage cost parameters can be interpreted as the disutility incurred from an increase in inventory of 100 ounces, or equivalently, the disutility of holding a certain number of 100 ounce bottles. For example, for the median consumer the disutility from holding 3 100 ounce bottles would be 0.0153, which is very small compared to the disutility of a one dollar increase in price, or of stocking out. The distribution of discount factors is qualitatively similar to the main results, with a significant spread and an upper quartile of close to 1. The average discount factor is somewhat lower at about 0.55. The main other difference between this specification and the results presented above is the lower stockout cost estimate. It is not surprising that this parameter is smaller, because we impose a tighter bound on how much people can store. If people sometimes hold inventories above our bound, the model will rationalize that behaviour with a higher stockout cost parameter. Finally, it is notable that the specification with increasing storage costs provides a much worse fit to the data (the deviance information criterion of this model is 44091.65, versus the estimate of 40510 for the main specification). This again is likely a result of our restrictions related

[24]Technically, we could allow a separate parameter for every possible value that storage costs could take, which would depend on the number of bottle combinations that an individual could hold. As we discuss in the text, we allow for 23 different bottle compositions. Adding an additional 23 parameters, with unobserved heterogeneity, would likely be infeasible and raise concerns about overfitting.

to inventory and that individuals treat the 80, 100 and 128 ounce sizes as the same, when these sizes have different shares in the data.

6 Counterfactuals

In our counterfactual exercises, we quantify the extent to which the model predictions can go wrong if the rational expectations assumption is incorrectly maintained. We assess two counterfactual exercises which vary the time series of price promotions: One is to increase the promotional depth, and the other is to increase the frequency of promotions.[25] Both of these counterfactuals will affect consumer stockpiling behavior, by inducing forward-looking individuals to stockpile more. We compute counterfactual quantities and revenues given a counterfactual price series by re-running the IJC procedure at the saved parameter draws for each estimation run. In particular, for a given draw, we simulate purchases, and then update the value function. This updated value function is saved for the next parameter draw, where we simulate choices given the next draw, and so on. Computing the counterfactuals in this way allows us to put confidence bounds around them (Ching, Imai, Ishihara, and Jain 2012).

In total, we compute simulated choices under six scenarios. The first three are done using the main model estimates, where we estimated individual-specific discount factors. Here, we simulate choices at i) the prices observed in the data, ii) given a larger promotional depth for the 100 oz Tide product, and iii) given a higher promotional frequency for the 100 oz Tide products. We also compute the same three counterfactuals using our estimates from the model specification with the discount factor fixed at the rational expectations benchmark. To increase promotional frequency, our procedure is to first identify deals in the store data using IRI's supplied price reduction flag. If the number of observed deals for a particular store-upc-year combination is N, we then double the number of promotions for that particular store-up-year by taking all non-deal prices, and randomly assigning N of those observations to be deals. To construct the price at the new deal observations, we multiply the observed regular retail price by the average discount observed in that particular store-upc-year. To increase promotional depth, we identify all promotional prices for our focal UPCs, and then divide the prices by 2.

The simulated changes in quantities and revenues for 100 oz Tide are shown in Table 6. At the estimated parameters, when promotions are deepened, the number of bottles sold rises by 502 units. In contrast, if one assumes rational expectations, one significantly overpredicts the effect

[25] Both of these counterfactual exercises were examined in Osborne (2018b) in the context of stockpiling, under the assumption of a fixed discount factor.

Table 6: Counterfactual Effect of Changes in Promotional Process on Quantities and Revenues

	Estimated Discount Factor		Rational Expectations	
Counterfactual	Quantity	Revenue	Quantity	Revenue
Increased Depth	502.72	1018.48	576.95	1183.98
	[448, 560]	[777.49, 1256.67]	[514, 641]	[916.38, 1452.15]
Increased Frequency	49.34	232.99	55.19	276.34
	[-11, 110]	[-214.35, 679.05]	[-7, 118]	[-171.78, 734.91]

Notes: This table shows the average estimated change in number of units sold (quantity) and revenues for each of the counterfactual price processes in the rows, compared to simulated choices from the prices observed in the data. The results are shown for the 100 OZ bottle of Tide. 95% confidence bounds are shown in brackets.

of deeper promotions on units sold. This is intuitive: When price promotions are made better, it is more worthwhile for forward-looking individuals to wait for them, rather than purchasing a different product or purchasing at a higher price. The counterfactual exercise also results in an overprediction of the revenue increase by about 10%, although the two estimated effects lie within each others' confidence bounds. In contrast, the impact of an increase in promotional frequency is relatively insensitive to the rational expectations assumption. Part of the reason for this is that overall, individuals seem less response to increases in promotional frequency than promotional depth (this finding is replicated in a different product category in Osborne (2018b), under the assumption of a fixed discount factor). In summary, using the rational expectations benchmark when it is inappropriate can generate incorrect forecasts, particularly if the objective is to forecast the impact of better promotions.

7 Conclusion

Consumer stockpiling behavior in packaged goods categories is often cited as an example of a situation where consumers are forward-looking. However, previous research (most notably, Erdem, Imai, and Keane (2003), Hendel and Nevo (2006a)) assumes (i) consumer are homogeneous in their discount factors, and (ii) consumers do not arbitrage and hence discount factor can be set according to the prevailing interest rate. Moreover, previous research has imposed smoothness assumptions on the storage cost function, assuming it to be continuous in inventory. We emphasize that this seemingly innocuous simplifying assumption rules out exclusion restrictions which naturally arise from the institutional features of this problem, which is why previous work fixed the discount factor rather than estimating it. By properly modelling storage cost as a step function of inventory (because storage cost only depends on the number of packages stored), the key state variable of this model, inventory, provides *natural exclusion restrictions* that can help identify the parameters of this model, including the discount factor. In our estimation results based on field data for laundry detergent, weekly discount factors average at around 0.71, lower than the value of 0.9995 this is obtained if one uses a common interest rate to set it. These differences are large, and are managerially relevant: if a brand manager were to assume individuals were very forward-looking, she would significantly overpredict the impact of offering better discounts on quantity sold. In our application, increasing promotional depth for a product drives more brand switching the more forward-looking individuals are. However, since individuals use laundry detergent at a constant rate and probably do not stock out for long periods of time, if a store manager were to offer better discounts for all brands at once, it is unlikely that a significant increase in category sales would arise.

There are many other avenues for future research on the relevance of forward-looking behavior. Although we demonstrate that exclusion restrictions can hold for storable product categories like laundry detergent, there may be ways to generate such restrictions in product categories where consumption is endogenous. Additionally, forward-looking behavior is an important driver of durable goods purchases, and estimating discount factors in those categories may generate additional insights.

References

Abbring, J. and O. Daljord (2018). Identifying the discount factor in dynamic discrete choice models. Working Paper.

Akça, S. and T. Otter (2015). Identifying the discount factor of forward looking consumers based on consumption from inventory. Working Paper.

Bronnenberg, B., M. Kruger, and C. Mela (2008). The IRI marketing data set. Marketing Science 27(4), 745–748.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. Econometric Theory 13(2), 185–213.

Chan, T., C. Narasimhan, and Q. Zhang (2008). Decomposing purchase elasticity with a dynamic structural model of flexible consumption. Journal of Marketing Research 45(4), 487–498.

Chevalier, J. and A. Goolsbee (2009). Are durable-goods consumers forward-looking? evidence from college textbooks. Quarterly Journal of Economics 124(4), 1854–1884.

Ching, A., T. Erdem, and M. P. Keane (2014). A simple method to estimate the roles of learning, inventories, and category consideration in consumer choice. Journal of Choice Modelling 13, 60–72.

Ching, A., S. Imai, M. Ishihara, and N. Jain (2012). A practitioner's guide to bayesian estimation of discrete choice dynamic programming models. Quantitative Marketing and Economics 10(2), 151–196.

Ching, A. and M. Ishihara (2018). Identification of dynamic models of rewards program.

Chung, D., T. Steenburgh, and K. Sudhir (2013). Do bonuses enhance sales productivity? a dynamic structural analysis of bonus-based compensation plans. Marketing Science 33(2), 165–187.

Coase, R. H. (1972). Durability and monopoly. Journal of Law and Economics 15, 143–149.

Crawford, G. and M. Shum (2005, July). Uncertainty and learning in pharmaceutical demand. Econometrica 37, 1137–1174.

Dubé, J.-P., G. Hitsch, and P. Jindal (2014, December). The joint identification of utility and discount functions from stated choice data: An application to durable goods adoption. Quantitative Marketing and Economics 12(4), 331–377.

Erdem, T., S. Imai, and M. Keane (2003). A model of consumer brand and quantity choice dynamics under price uncertainty. Quantitative Marketing and Economics 1(1), 5–64.

Erdem, T. and M. Keane (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. Marketing Science 15(1), 1–20.

Fang, H. and Y. Wang (2015). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. International Economic Review.

Frederick, S., G. Loewenstein, and T. O'Donoghue (2002). Time discounting and time preference: A critical review. Journal of Eocnomic Literature 15, 351–401.

Geweke, J. and M. Keane (2000). Bayesian inference for dynamic choice models without the need for dynamic programming. Cambridge University Press. in Roberto Mariano, Til Schuermann, Melvyn J. Weeks (eds.), *Simulation-based Inference in Econometrics: Methods and Applications*.

Hartmann, W. R. and H. S. Nair (2010). Retail competition and the dynamics of demand for tied goods. Marketing Science 29(2), 366–386.

Haviv, A. (2014). Does purchase without search explain counter cyclic pricing? working paper, Simon Business School, University of Rochester.

Hendel, I. and A. Nevo (2006a). Measuring the implications of sales and consumer inventory behavior. Econometrica 74, 1637–1673.

Hendel, I. and A. Nevo (2006b). Sales and consumer inventory. The RAND Journal of Economics 37, 543–561.

Hendel, I. and A. Nevo (2013). Intertemporal price discrimination in storable goods markets. American Economic Review 103(7), 2722–2751.

Hong, P., P. McAfee, and A. Nayyar (2002). Equilibrium price dispersion with consumer inventories. Journal of Economic Theory 105, 503–517.

Hotz, J. and R. Miller (1993, July). Conditional choice probabilities and the estimation of dynamic models. The Review of Economic Studies 60(3), 497–529.

Houser, D., M. Keane, and K. McCabe (2004). Behavior in a dynamic decision problem: An analysis of experimental evidence using a bayesian type classification algorithm. Econometrica 72, 781–822.

Imai, S., N. Jain, and A. Ching (2009). Bayesian estimation of dynamic discrete choice models. Econometrica 77(6), 1865–1899.

Ishihara, M. and A. Ching (2012). Dynamic demand for new and used durable goods without physical depreciation: The case of japanese video games. Working Paper, available at: https://ssrn.com/abstract=2189871.

Lee, R. S. (2013). Vertical integration and exclusivity in platform and two-sided markets. American Economic Review 103(6), 2960–3000.

Liu, Y. and S. Balachander (2014). How long has it been since the last deal? consumer promotion timing expectations and promotional response. Quantitative Marketing and Economics 12(1), 85–128.

Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete decision processes. Econometrica 20(2), 801–816.

McFadden, D. (1981). Econometric Models of Probabilistic Choice. MIT Press: Cambridge, MA. in C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Application*.

Osborne, M. (2011, March). Consumer learning, switching costs and heterogeneity: A structural examination. Quantitative Marketing and Economics 9(1), 25–70.

Osborne, M. (2018a). Approximating the cost-of-living index for a storable good. American Economic Journal: Microeconomics 10(2), 286–314.

Osborne, M. (2018b). Frequency versus depth: How changing the temporal process of promotions impacts demand for a storable good. Japanese Economic Review 69(3), 258–283.

Pesendorfer, M. (2002). Retail sales: A study of pricing behavior in supermarkets. Journal of Business 75(1), 33–66.

Pires, T. (2016). Costly search and consideration set in storable goods markets. Quantitative Marketing and Economics 14(3), 157–193.

Rossi, F. (2017). Lower price or higher reward? measuring the effect of consumers' preferences on reward programs. Forthcoming in Management Science.

Rossi, P., G. Allenby, and R. McCulloch (2005). Bayesian Statistics and Marketing. Wiley.

Rust, J. (1994). Structural Estimation of Markov Decision Processes. Elsevier. in Engle R. and McFadden D (eds.), *Handbook of Econometrics*.

Ruud, P. A. (1996). Simulation of the multinomial probit model: An analysis of covariance matrix estimation. Working paper.

Seiler, S. (2013). The impact of search costs on consumer behavior: A dynamic approach. Quantitative Marketing and Economics 11(2), 155–203.

Sobel, J. (1984). The timing of sales. Review of Economic Studies 51, 353–368.

Sun, B. (2005). Promotion effect on endogenous consumption. Marketing Science 24, 430–443.

Yang, B. and A. Ching (2014). Dynamics of consumer adoption of financial innovation: The case of atm cards. Management Science 60(4), 903–922.

Yao, S., C. Mela, J. Chiang, and Y. Chen (2012). Determining consumers' discount rates with field studies. Journal of Marketing Research 49(6), 822–841.

ONLINE APPENDICES [Intended to be made available online.]

A Derivations of Parameters from Choice Probabilities

This section presents the derivations used in the constructive proof of identification presented in Section 4. We use the notation from that section that $v_j(I)$ represents the choice-specific value for option j at inventory level I. To start, recall that:

$$v_1(I) = -\alpha p - \omega_{B(I)+1} + \beta V(I + b - 1), \tag{16}$$

$$v_0(I) = -\omega_{B(I)} - \nu \mathbf{1}\{I = 0\} + \beta V(I - 1), \tag{17}$$

where $B(I)$ is the number of bottles held at inventory level I, and $\omega_0 = 0$.

We are interested in considering the difference in choice-specific value for these two options because they are simply a function choice probabilities by Hotz and Miller (1993)'s Inversion Theorem. For example, if an individual's inventory I is in the range $[2, ..., b + 1]$, then the difference in choice-specific values can be written as

$$v_1(I) - v_0(I) = -\alpha p - (\omega_2 - \omega_1) + \beta[V(I + b - 1) - V(I - 1)] \tag{18}$$

$$\Delta v(I) = -\alpha p - \Delta\omega(2, 1) + \beta[V(I + b - 1) - V(I - 1)], \tag{19}$$

where $\Delta v(I) \equiv v_1(I) - v_0(I) = \ln(P_1(I)/P_0(I))$; $\Delta\omega(2, 1) \equiv \omega_2 - \omega_1$. $\ln(P_1(I)/P_0(I))$ are the observed choice probabilities at inventory level I.

Note that our exclusion restriction assumption implies that if we take the difference of $\Delta v(.)$ at two different values of I that satisfy the exclusion restriction requirements, we can difference out the current utility components.

$$\begin{aligned} \ln(P_1(I + 1)/P_0(I + 1)) - \ln(P_1(I)/P_0(I)) &= \Delta v(I + 1) - \Delta v(I) \\ &= \beta[V(I + b) - V(I) - (V(I + b - 1) - V(I - 1))]. \end{aligned} \tag{20}$$

Note that if the LHS of (20) is not zero for any value of I where the exclusion restriction holds ($I > 1$, and values of I not on the boundaries where storage costs change), that implies $\beta > 0$ by contraposition. If the LHS of (20) is zero for all values of I where the exclusion restriction holds, then it is clear that $\beta = 0$ rationalizes the observed choice probabilities. We note that in practice, such a situation is a measure zero event and will never occur. Thus, in the derivation of β below we assume $\beta > 0$. We use the definition of Emax and $\ln\left(\frac{P_1(I)}{P_0(I)}\right) = \Delta v(I)$ (see Ching and Ishihara (2018)) to derive,

$$v_0(I) = V(I) - \ln\left(1 + \frac{P_1(I)}{P_0(I)}\right). \tag{21}$$

Suppose WLOG that an individual has a single package in inventory. Then we can write,

$$
\begin{aligned}
V(I-1) &= \frac{1}{\beta}\left(v_0(I)+\omega_1\right) \\
&= \frac{1}{\beta}\left(V(I)-\ln\left(1+\frac{P_1(I)}{P_0(I)}\right)+\omega_1\right), \qquad (22)
\end{aligned}
$$

where the second equality follows from Eq(21). Then we know that

$$
\begin{aligned}
V(I)-V(I-1) &= V(I)-\frac{1}{\beta}\left(V(I)-\ln\left(1+\frac{P_1(I)}{P_0(I)}\right)-\omega_1\right) \\
&= \left(1-\frac{1}{\beta}\right)V(I)+\frac{1}{\beta}\left(\ln\left(1+\frac{P_1(I)}{P_0(I)}\right)+\omega_1\right).
\end{aligned}
$$

Then using this expression for $V(I)-V(I-1)$, we can rewrite Eq(20) as

$$
\begin{aligned}
\Delta v(I+1)-\Delta v(I) &= \beta\Bigg[\left(1-\frac{1}{\beta}\right)V(I+b)+\frac{1}{\beta}\left(\ln\left(1+\frac{P_1(I+b)}{P_0(I+b)}\right)+\omega_2\right)- \\
&\quad \left(\left(1-\frac{1}{\beta}\right)V(I)+\frac{1}{\beta}\left(\ln\left(1+\frac{P_1(I)}{P_0(I)}\right)+\omega_1\right)\right)\Bigg] \\
&= (\beta-1)(V(I+b)-V(I))+\Phi(I)+\Delta\omega(2,1), \qquad (23)
\end{aligned}
$$

where

$$
\Phi(I)=\ln\left(1+\frac{P_1(I+b)}{P_0(I+b)}\right)-\ln\left(1+\frac{P_1(I)}{P_0(I)}\right).
$$

Note that by the definition of v_0, v_1,

$$
V(I+b)-V(I)=\frac{1}{\beta}\left(\Delta\omega(2,1)+\alpha p+v_1(I+1)-v_0(I+1)\right).
$$

Then plugging this into Eq(23), we get:

$$
\begin{aligned}
\Delta v(I+1)-\Delta v(I) &= (\beta-1)\left(\frac{1}{\beta}\left(\Delta\omega(2,1)+\alpha p+v_1(I+1)-v_0(I+1)\right)\right)+\Phi(I)+\Delta\omega(2,1) \\
&= \frac{\beta-1}{\beta}\left(v_1(I+1)-v_0(I+1)\right)+\Phi(I)+\frac{2\beta-1}{\beta}\Delta\omega(2,1)+\frac{\beta-1}{\beta}\alpha p.
\end{aligned}
$$

To simplify the notation, denote $\Delta v(I)=v_1(I)-v_0(I)$. Then we can write the above equation as:

$$
\Delta v(I+1)-\Delta v(I) = \frac{\beta-1}{\beta}\Delta v(I+1)+A(I)+\frac{2\beta-1}{\beta}\Delta\omega(2,1)+\frac{\beta-1}{\beta}\alpha p.
$$

We can simplify the above equation to,

$$
\Delta v(I)=\frac{1}{\beta}\Delta v(I+1)-\Phi(I)-\frac{2\beta-1}{\beta}\Delta\omega(2,1)+\frac{\beta-1}{\beta}\alpha p. \qquad (24)
$$

Note that we cannot solve Eq (24) for β, as it involves $\Delta\omega(2,1)$. Thus, consider solving the system of equations

$$\Delta v(I) = \frac{1}{\beta}\Delta v(I+1) - \Phi(I) - \frac{2\beta-1}{\beta}\Delta\omega(2,1) + \frac{\beta-1}{\beta}\alpha p. \tag{25}$$

$$\Delta v(I+1) = \frac{1}{\beta}\Delta v(I+2) - \Phi(I+1) - \frac{2\beta-1}{\beta}\Delta\omega(2,1) + \frac{\beta-1}{\beta}\alpha p, \tag{26}$$

where the exclusion restrictions hold for inventory levels I through $I+2$. To solve for β, we can subtract Eq(25) from Eq(26), which gives us the following equation:

$$\Delta v(I+1) - \Delta v(I) \quad = \quad \frac{1}{\beta}\left(\Delta v(I+2) - \Delta v(I+1)\right) - \left(\Phi(I+1) - \Phi(I)\right).$$

We can solve the above equation to express β as a function of choice probabilities:

$$\begin{aligned}\beta \quad &= \quad \frac{\Delta v(I+2) - \Delta v(I+1)}{\Delta v(I+1) - \Delta v(I) + \Phi(I+1) - \Phi(I)} \\ &= \quad \frac{\Delta \log(P(I+2)) - \Delta \log(P(I+1))}{\Delta \log(P(I+1)) - \Delta \log(P(I)) + \Phi(I+1) - \Phi(I)},\end{aligned} \tag{27}$$

where

$$\Delta \log(P(I)) \equiv \log(P_1(I)) - \log(P_0(I)) = \Delta v(I).$$

The Δv's are differences in log choice probabilities, and $\Phi(I)$ depends on choice probabilities at $I+b$ and I. Note that to achieve identification of β, the researcher must be able to estimate probabilities at five different inventory values: $I, I+1, I+2, I+b$, and $I+b+1$. For identification to be possible, it needs to be the case that individuals can hold at least two packages (i.e., $M \geq 2$), and that the denominator of (27) is nonzero (a rank condition).

The price coefficient, α, can identified from the purchase probability when $I = Mb$, the maximum inventory bound. At this inventory level purchasing does not increase storage cost or affect the value function. i.e.,

$$v_1(Mb) = -\alpha p - \omega_M + \beta V(Mb-1),$$

$$v_0(Mb) = -\omega_M + \beta V(Mb-1).$$

As a result, $\alpha = -\Delta v(Mb)/p$. As we discuss in the body of the paper, in general it is likely preferred to appeal to price variation to identify α.

Given a solution for β, and assuming that α is also identified, we can solve for ω_1 and ν. To do this, we use the following equations which are derived from choice probabilities at inventory levels of 0 and 1:

$$\Delta v(0) = -\omega_1 - \nu - \alpha p + \beta(V(b-1) - V(0)), \tag{28}$$

and

$$\Delta v(1) = -\omega_1 - \alpha p + \beta(V(b) - V(0)). \tag{29}$$

Below, we show how to express the differences in values functions in Eq(28) and Eq(29) in terms of parameters and choice probabilities. To start, we first derive the value function at $I = 0$:

$$\begin{aligned} V(0) &= v_0(0) + \ln\left(1 + \frac{P_1(0)}{P_0(0)}\right) \\ &= -\nu + \beta V(0) + \ln\left(1 + \frac{P_1(0)}{P_0(0)}\right). \end{aligned}$$

Thus the equation for $V(0)$ is

$$V(0) = \frac{-\nu}{1-\beta} + \frac{1}{1-\beta}\ln\left(1 + \frac{P_1(0)}{P_0(0)}\right).$$

Then we can write the equation for $V(1)$ recursively as

$$\begin{aligned} V(1) &= v_0(1) + \ln\left(1 + \frac{P_1(1)}{P_0(1)}\right) \\ &= \beta V(0) + \ln\left(1 + \frac{P_1(1)}{P_0(1)}\right) \\ &= \frac{-\beta\nu}{1-\beta} + \frac{\beta}{1-\beta}\ln\left(1 + \frac{P_1(0)}{P_0(0)}\right) + \ln\left(1 + \frac{P_1(1)}{P_0(1)}\right). \end{aligned}$$

and doing a similar manipulation with $V(2)$ we have

$$V(2) = -\omega_1 - \frac{\beta^2\nu}{1-\beta} + \frac{\beta^2}{1-\beta}\ln\left(1 + \frac{P_1(0)}{P_0(0)}\right) + \beta\ln\left(1 + \frac{P_1(1)}{P_0(1)}\right) + \ln\left(1 + \frac{P_1(2)}{P_0(2)}\right).$$

In general, for $2 \leq I \leq b+1$, we can write:

$$\begin{aligned} V(I) &= -\sum_{i=0}^{I-2}\beta^i\omega_1 - \frac{\beta^I\nu}{1-\beta} + \frac{\beta^I}{1-\beta}\ln\left(1 + \frac{P_1(0)}{P_0(0)}\right) + \sum_{i=0}^{I-1}\beta^i\ln\left(1 + \frac{P_1(I-i)}{P_0(I-i)}\right) \\ &= -\frac{1-\beta^{I-1}}{1-\beta}\omega_1 - \frac{\beta^I\nu}{1-\beta} + g(\beta, I, P_0(0), P_1(0), ..., P_0(I), P_1(I)), \end{aligned}$$

where

$$g(\beta, I, P_0(0), P_1(0), ..., P_0(I), P_1(I)) = \frac{\beta^I}{1-\beta}\ln\left(1 + \frac{P_1(0)}{P_0(0)}\right) + \sum_{i=0}^{I-1}\beta^i\ln\left(1 + \frac{P_1(I-i)}{P_0(I-i)}\right). \tag{30}$$

Therefore we can now write the value function differences as:

$$V(b-1)-V(0)=-\frac{1-\beta^{b-2}}{1-\beta}\omega_1+\frac{1-\beta^{b-1}}{1-\beta}\nu+h_0(\beta),$$

$$V(b)-V(0)=-\frac{1-\beta^{b-1}}{1-\beta}\omega_1+\frac{1-\beta^{b}}{1-\beta}\nu+h_1(\beta),$$

where the h functions are defined as

$$h_0(\beta) = g(\beta, b-1, P_0(0), P_1(0), ..., P_0(I), P_1(I)) - \frac{1}{1-\beta}\ln\left(1+\frac{P_1(0)}{P_0(0)}\right), \tag{31}$$

$$h_1(\beta) = g(\beta, b, P_0(0), P_1(0), ..., P_0(I), P_1(I)) - \frac{1}{1-\beta}\ln\left(1+\frac{P_1(0)}{P_0(0)}\right). \tag{32}$$

Therefore the two equations Eq(28) and Eq(29) which define the choice-specific values at states 0 and 1 can be rewritten as

$$\Delta v(0)=\frac{\beta^{b-1}-1}{1-\beta}\omega_1+\frac{2\beta-\beta^{b}-1}{1-\beta}\nu-\alpha p+\beta h_0(\beta), \tag{33}$$

and

$$\Delta v(1)=\frac{\beta^{b}-1}{1-\beta}\omega_1+\frac{\beta-\beta^{b+1}}{1-\beta}\nu-\alpha p+\beta h_1(\beta). \tag{34}$$

The $\Delta v(I)$ above are equal to $\ln(P_1(I)/P_0(I))$, and the h functions depend only on $beta$ (which we have solved for in Eq((27)) and choice probabilities. Hence, the above defines a system of two equations in two unknowns which we can use to solve for ν and ω_1.

Given the solutions of the above parameters, we can recursively solve for the other storage cost parameters, ω_2 through ω_M, in terms of choice probabilities and known parameters. First, consider the parameter ω_2. We can solve directly for ω_2 using the choice probabilities at inventory states $I=2$ through $I=b+1$. The simplest formula for ω_2 arises when we use $I=2$, as the log difference in choice probabilities at this state is

$$\Delta v(2)=-\omega_2-\alpha p+\omega_1+\beta(V(b+1)-V(1)),$$

and we have already solved for the parameters β, α, as well as the two value functions in terms of choice probabilities, ω_1, β, and α (if we used an inventory state higher than 2, the value function difference above would depend on ω_2 as well). Using derivations from above the value function difference can be expressed in terms of choice probabilities and known parameters:

$$\begin{aligned} V(b+1)-V(1) &= -\frac{1-\beta^{b}}{1-\beta}\omega_1-\frac{\beta^{b}}{1-\beta}\nu+g(\beta, b+1, P_0(0), ..., P_1(b+1)) \\ &\quad +\frac{\beta\nu}{1-\beta}-\frac{\beta}{1-\beta}\ln\left(1+\frac{P_1(0)}{P_0(0)}\right)-\ln\left(1+\frac{P_1(1)}{P_0(1)}\right), \end{aligned}$$

and one can express ω_2 as

$$\omega_2 = -\alpha p + \omega_1 + \beta(V(b+1) - V(1)) - (\ln(P_1(2)) - \ln(P_0(2)).$$

We can solve for higher levels of the storage costs, i.e., ω_3, ω_4, etc, using the choice probabilities at the inventory states $I = b+2, 2b+2, ...$ For example, consider solving for ω_3 at $I = b+2$. The relevant difference in choice probabilities is

$$\Delta v(b+2) = -\omega_3 + \omega_2 - \alpha p + \beta(V(2b+1) - V(b+1)).$$

As was the case with the solution for ω_2, we need an expression for the value function at $2b+1 \geq I \geq b+2$, which is

$$\begin{aligned} V(I) &= -\sum_{i=0}^{I-(b+2)} \beta^i \omega_2 - \sum_{i=I-(b+1)}^{I-2} \beta^i \omega_1 - \frac{\beta^I \nu}{1-\beta} + \frac{\beta^I}{1-\beta} \ln\left(1 + \frac{P_1(0)}{P_0(0)}\right) + \sum_{i=0}^{I-1} \beta^i \ln\left(1 + \frac{P_1(I-i)}{P_0(I-i)}\right) \\ &= -\frac{1-\beta^{I-(b+1)}}{1-\beta}\omega_2 - \beta^{I-(b+1)}\frac{1-\beta^b}{1-\beta}\omega_1 - \frac{\beta^I \nu}{1-\beta} + g(\beta, I, P_0(0), P_1(0), ..., P_0(I), P_1(I)). \end{aligned}$$

and it is straightforward to solve for ω_3. The logic above can be extended to ω_4, and higher values of storage costs.

We should note that the proof here is much more involved than the one in Ching and Ishihara (2018), who consider a simple dynamic store choice problem with rewards programs. They are able to take advantage of the feature that the value of state variables (rewards points) remain unchanged when one chooses the outside option. This feature allows them to significantly simplify the proof, but it is not available in our model. In particular, unlike our proof, they did not make use of the difference-in-difference (with respect to the choice specific value functions) restrictions at all.[26]

We provide a summary of which moment conditions we use to identify each parameter in Table 7. The parameter α is identified from the choice probability at inventory level $I = Mb$. To identify β, we exploit two moment conditions, which are the differences in the log ratio of choice probabilities at two consecutive inventory levels. These inventory levels must be chosen where the exclusion restrictions hold, and are chosen so they do not overlap with any of the moment conditions to use any other parameters (hence the range of inventory values are $\{3, \cdots, b-1\}$. For example, one could use the moment conditions $\ln(P_1(4)/P_0(4)) - \ln(P_1(3)/P_0(3))$ and $\ln(P_1(5)/P_0(5)) - \ln(P_1(4)/P_0(4))$.

[26] Rossi (2017) applies the identification results in Ching and Ishihara (2018) to estimate consumer's discount factor using the data from a retail gasoline reward program.

Table 7: Summary of Moment Conditions Used to Identify Model Parameters

Parameter(s)	Moment Conditions for Identification
α	$\ln(P_1(Mb)/P_0(Mb))$
β	$\ln(P_1(I+1)/P_0(I+1)) - \ln(P_1(I)/P_0(I))$,
	and $\ln(P_1(I+2)/P_0(I+2)) - \ln(P_1(I+1)/P_0(I+1))$, for $I \in \{3, \cdots, b-1\}$
ω_1, ν	$\ln(P_1(0)/P_0(0))$ and $\ln(P_1(1)/P_0(1))$
ω_2	$\ln(P_1(2)/P_0(2))$
ω_3	$\ln(P_1(b+2)/P_0(b+2))$
$\vdots$	$\vdots$
ω_M	$\ln(P_1((M-1)b+2)/P_0((M-1)b+2))$

The parameters ν and ω_1 are solutions to the moment equations relating choice-specific values to choice probabilities at $I = 0$ and $I = 1$. Finally, higher values of ω_B can be derived from the moment conditions defining $\ln(P_1((B-1)b+2)/P_0((B-1)b+2))$.

It is notable that in order to point identify β, we require two moment conditions - in other words, the exclusion restriction needs to hold for at least 3 values of inventory. The fact that a single exclusion restriction may not be sufficient for point identification of β has been shown by Abbring and Daljord (2018). In particular, they show that Eq ((20)) can have multiple (but finitely many) solutions for β. We provide an illustrative example of how in our application a single exclusion restriction may not allow point identification, while two exclusion restrictions can, in Figure 3. In the top panel, we plot the difference in value functions from the right hand side of equation (20) as a function of β for $I = 3$. If the researcher observes that $\Delta v(4) - \Delta v(3) = -0.14$, there are two possible values of β that can fit this moment, at 0.79 and 0.95. In the bottom panel of Figure 3, we plot the right hand side of equations (20) for two consecutive values of I, at 3 and 4. Now, if $\Delta v(4) - \Delta v(3) = -0.14$, and $\Delta v(5) - \Delta v(4) = -0.10$, there is only a single value of β that can rationalize these two moments, at 0.79. We note that we do not need more than two exclusion restrictions, as Eq ((27)) provides a formula for β based on two moments.

B Imputed Consumption Rates

In this section, we verify the realism of our assumptions that consumption rates are constant within an individual, and can be calibrated by dividing total quantity purchased by total number

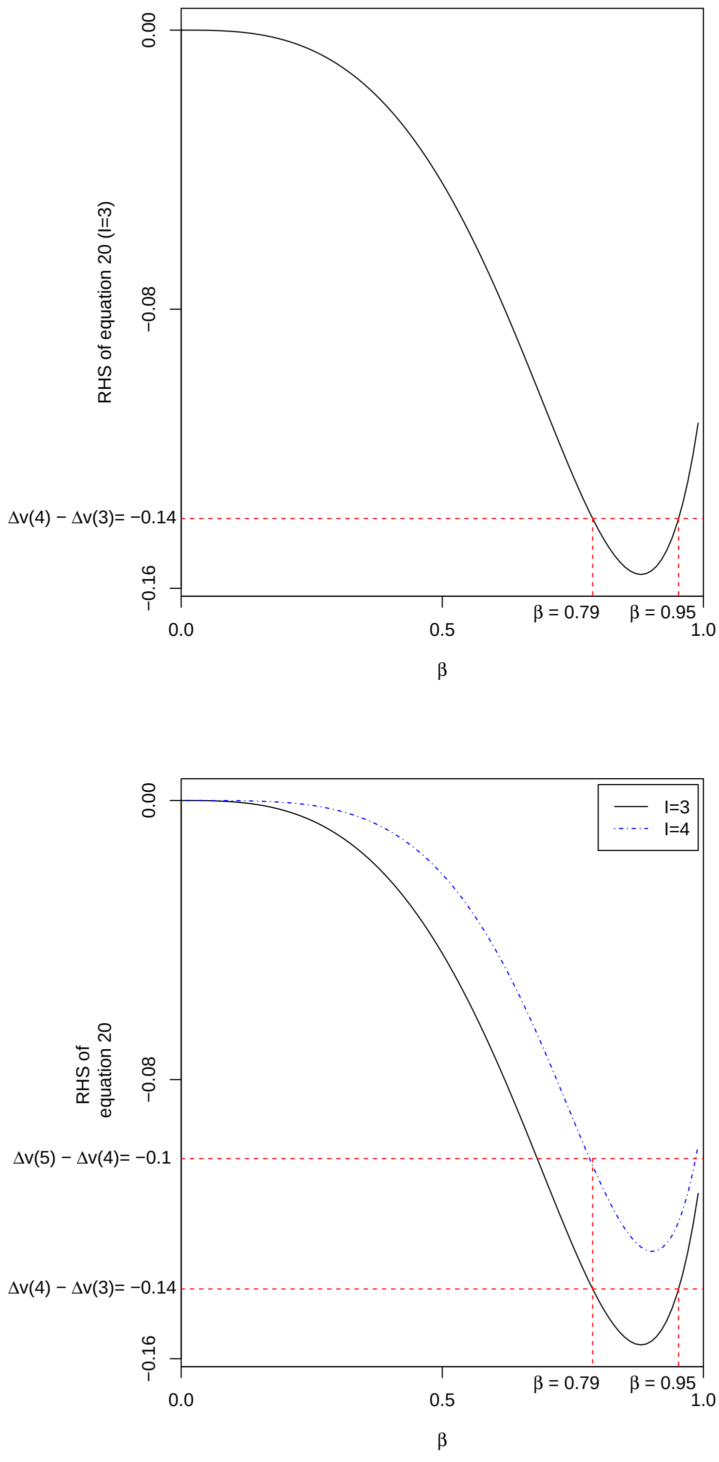

Figure 3: Plots of the right hand side of equation (20), $\beta(V(I+b)-V(I)-(V(I+b-1)-V(I-1))$, and possible solutions for β. Top panel: plot at $I=3$, with two possible solution for β. Bottom Panel, plots at $I=3$ and $I=4$, with only a single possible solution for β. Parameter values are $\nu=0.4$, $\eta=1$,$M=3$,$p=3.31$, $b=8$ and logit error term.

of weeks. Regarding the consumption rate calibration, one issue we were concerned about was that since our sample period (both pre-estimation and estimation sample) spans 7 years, and consumption rates might change over time due to external factors such as changes in household composition. If such changes occur, in our estimation we would want to use a consumption rate that reflected consumption in the estimation period, rather than the overall 7 year period. However, if consumption rates do not change, it would be preferable to use the entire period since the estimated consumption rate will be more precise. To address this potential issue, we perform the following exercise:

1. Compute consumption rates using the entire sample period.

2. Compare average, within household inventory in the pre-estimation to the estimation period.

3. For households where the difference between pre-estimation and estimation period inventory is more than 100 ounces (1 standard size bottle), use the consumption rate calculated from the estimation period purchases only.

In our estimation sample, we use the post-estimation period to infer consumption rates for 71% of the sample. We present the distribution of implied consumption rates and inventories for the estimation sample in Table 8. In the empirical model, we estimate the consumption rate in terms of the number of ounces a household would use per week. In the table, we present the consumption rates in terms of the number of loads of laundry a household would do per week. To arrive at this number, we use the rule that 100 ounces of laundry detergent converts to 32 loads of laundry. The table suggests that households are doing 3 to 4 loads per week. Imputed inventory is also measured in ounces. We divide this number by 100 to convert it into 100 ounce bottle equivalents, and present the distribution of within household average inventory in the second row of the table. Our estimated consumption rates suggest that individuals hold between 3 and 4 100 ounce bottles at a time. Overall, the imputed consumption rates and inventory distributions seem reasonable.

We perform two additional exercises to assess the reasonableness of our imputed consumption rates. First, we regress the consumption rates on household demographics that we might expect to be correlated with them. The results of this regression are shown in Table 9. The two most significant coefficients are household size of 4 people, and whether an individual has 2 dogs, both of which should be correlated with higher consumption rates. The second exercise we perform is to regress imputed inventory in week $t-1$ on a dummy variable for purchase in week t. Table 10 shows the results of a spline regression of a dummy for purchase on different levels of imputed inventory. When constructing the independent variables we classify an individual's inventory into

Table 8: Distribution of Consumption Rates And Inventories

Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Consumption Rates (Loads)	1.22	2.87	3.95	4.37	5.54	13.07
Imputed Inventory (100 Oz Bottles, overall)	0	1.5	3.1	3.97	5.78	14.98
Imputed Inventory (100 Oz Bottles, within household)	0.37	1.76	2.97	3.94	5.37	12.71

Notes: These statistics are constructed from the 312 households in the estimation data. The second row shows the distribution across households of the average within-household imputed inventory during the estimation period.

three categories: above 0 and less than 50 ounces (low), between 50 and 100 ounces (medium), and between 100 and 200 ounces (high). These cutoffs correspond to three popular bottle sizes in the data. The excluded category is any inventory level above 200 ounces. An inventory level of 0 is classified as a stockout. In Table 10, $0 < Inv \leq 50$, $50 < Inv \leq 100$ and $100 < Inv \leq 200$ are dummy variables for each of these inventory levels, while $0 < Inv \leq 50 \times$ Inv, etc, are interactions between the dummy variable and the inventory level. The coefficient estimates are consistent with how individuals should behave if they stockpile and are forward-looking. In particular, purchase probabilities are highest if a stockout occurs, and they decrease as inventory rises. Moreover, the slope of the purchase probability with respect to the inventory level drops as inventory rises, which we use to help verify rank condition R1.

C Extension to Inclusive Value Sufficiency

This section describes in detail how the extensions to Inclusive Value Sufficiency (IVS) first proposed in Osborne (2018a) can be used to reduce the size of the price state space. Recall that we index brands by k, package sizes by x, and the number of packages chosen by j. Under standard IVS, where error terms are assumed to follow a logit distribution and brand level utilities do not scale with the number of packages chosen, the inclusive value will be a function of both the package size chosen, x, and the number of packages chosen, j:

$$\Omega_{it}(x,j) = \ln\left(\sum_k \exp(\xi_{ixk} - \alpha_i p_{ixkt} j)\right).$$

The number of inclusive values one would have to track will be equal to XK. The idea behind the extension to IVS is to essentially be able to factor j out of the inclusive value. To do this, two

Table 9: Regression of Consumption Rate on Demographics

Regressor	Estimate
HH Size 2	0.147*
HH Size 3	0.189**
HH Size 4	0.3478***
HH Size 5	0.1499
HH Size 6	0.266
Homeowner	-0.0757
Child 12-17	0.1189
Children 0-5 and 6-11	-0.013
Children 6-11 and 12-17	-0.0584
Children 0-5, 6-11, and 12-17	0.3715
No Children	-0.0917
Num of Dogs 1	0.0666
Num of Dogs 2	0.3195***
Num of Dogs 3	-0.3286
Num of Cats 1	0.0288
Num of Cats 2	0.0123
Num of Cats 3	-0.0061
Num of Cats 4	-0.0126
Num of Cats 5	0.1892
R Squared	0.22

Note: This regression is based on 312 observations.

Table 10: Regression of Purchase Dummy on Inventory Level

Regressor	Estimate
Intercept	0.0857***
Stockout	0.1338***
$0 < \text{Inv} \leq 50$	0.1295***
$0 < \text{Inv} \leq 50 \times \text{Inv}$	-0.1279**
$50 < \text{Inv} \leq 100$	0.0801**
$50 < \text{Inv} \leq 100 \times \text{Inv}$	-0.0516
$100 < \text{Inv} \leq 200$	0.0751***
$100 < \text{Inv} \leq 200 \times \text{Inv}$	-0.0362***
R Squared	0.0117

Notes: This regression is based on 35515 household-week observations. The inventory levels reflect inventories in the prior week, measured in hundreds of ounces.

assumptions are necessary. First, that brand level utility scales with j, and second, that the error term follows a nested logit distribution where the outer nest corresponds to the choice of package size and number of packages, and the inner nest corresponds to the brand choice. Regarding the first assumption, we assume that the flow utility for choosing j packages of brand k can be written as $\frac{j}{J}\xi_{ixk}$, so that the inclusive value can be written as

$$\Omega_{it}(x,j) = \ln\left(\sum_k \exp\left(\frac{j}{J}[\xi_{ixk} - \alpha_i J p_{ixkt}]\right)\right).$$

Formally, the second assumption means that the choice-specific error can be decomposed into two errors as follows:

$$\varepsilon_{ijxkt} = e_{ijxt} + \frac{j}{J}v_{ikt},$$

where the distribution of v_{ikt} is Type 1 extreme value and the distribution of e_{ijxt} has a distribution of the form denoted as $C(\lambda)$ from Cardell (1997), where $\lambda = j/J$ (note the $C(\lambda)$ notation is introduced in Cardell (1997) to refer to a particular distribution, and should not be confused with our notation for the consumer's choice set). It can be shown using derivations from McFadden (1981) that if a choice-specific error follows an extreme value distribution with scale parameter λ then the expected utility can be written as

$$EU = \lambda \ln \left(\sum_k \exp\left(\frac{j}{J} [\xi_{ixk} - \alpha_i J p_{ixkt}] / \lambda \right) \right).$$

Under the assumption that the choice-specific error across package sizes, number of packages and brands follows a nested logit distribution, we need the λ parameter to be between 0 and 1, which is why we assume that it is equal to j/J. As a result, the inclusive value can be written as

$$\Omega_{it}(x, j) = \frac{j}{J} \ln \left(\sum_k \exp(\xi_{ixk} - \alpha_i J p_{ixkt}) \right).$$

Note that since j can be factored out of the inclusive value, it is not necessary to track a different inclusive value for each j, and it is sufficient to write the inclusive value as $\Omega_{it}(x)$.

D Model with Increasing Storage Costs

This section describes an extended version of the empirical model in the main text, where we allow for storage costs to increase as the number of bottles held increases. As we discuss in the body of the paper, there are three issues that arise when allowing for increasing storage costs in markets where different package sizes of a product are available. Two of these are computational: it becomes necessary to track the composition of inventory, which substantially increases the size of the model's state space, and if one wishes to use a flexible functional form to estimate the storage cost function, the number of parameters one needs to include rises substantially. A third issue is conceptual: one needs to make an assumption about the order in which consumption from inventory occurs. To address the computational concerns, we simplify the state space by allowing individuals to only hold up to 6 bottles, and by grouping the three medium sized bottles together, and we reduce the number of parameters needed by assuming quadratic storage costs. To address the conceptual concerns, we assume that the order of consumption in inventory is first-in-first-out (FIFO), meaning that the order of consumption corresponds to the order of purchase. While it is difficult to verify the validity of the FIFO assumption, Akça and Otter (2015) provide survey evidence for four product categories (they do not include laundry detergent, but do include dishwashing detergent) that individuals tend to only keep one package of a product open at a time.

Operationally, we assume that an individual can hold up to $N_B = 6$ bottles, and keep track of the size of bottle that is stored in slot $o = 1, ..., N_B$. The FIFO assumption means that the bottle in slot 1 is currently being consumed, and once it is finished, the bottle in slot 2 will be moved to slot 1, the one in 3 will be moved to 2, and so on. In this version of the model we keep track of 2 inventory states, one continuous and one discrete. The continuous state is the amount of inventory left in bottle 1 in ounces, which we denote as I. The discrete state is the current configuration of bottles, $\boldsymbol{B} = (b_1, b_2, ..., b_{N_B})$. We assume that storage the storage cost function is a quadratic function of the total volume taken up by the current configuration of bottles $\boldsymbol{B}$:

$$s(\boldsymbol{B}; \boldsymbol{\omega}_i) = \omega_{i1} \left(\sum_{o=1}^{N_B} Vol(b_o) \right) + \omega_{i2} \left(\sum_{o=1}^{N_B} Vol(b_o) \right)^2,$$

where $Vol(b_o)$ is the volume of the bottle in slot o. Even though the bottle composition state is discrete, the number of states we have to track can get potentially very large in N_B is big. For example, if $N_B = 6$, and the number of package sizes is 5, the total number of possible bottle composition states is 19,531. We discretize the continuous part of the inventory state (the amount left in the currently open bottle) into 17 points, and we also track 100 price states, which means for each household we would have to track 33,202,700 states. It is clearly computationally infeasible to include this many states per household. As a result, we make two simplifications. First, recall that there are five package sizes individuals buy: 50 ounces, 80 ounces, 100 ounces, 128 ounces, and 200 ounces. We group the 80, 100 and 128 ounce sizes together, meaning that we assume that individuals treat these sizes as though they contain 100 ounces of detergent. One way to interpret this assumption is that households perceive these sizes as allowing them to do the same number of loads of laundry, which is probably not unrealistic: If one uses the rule that 100 ounces produces 32 loads, then the 80 ounce bottle produces 26 loads, and the 128 ounce bottle produces 40. Since it takes households several weeks to use up a bottle of detergent, it is probably difficult for households to predict exactly how long a bottle will last. Moreover, the volume taken up by all these households will likely be similar.

Even if we assume that individuals only hold 3 package sizes, that leaves us with 781 discrete states, which is still large. A second simplification we make is to reduce the set of admissable states to those that are most frequently visited in the data. Note that under the assumption of constant consumption rates which are estimated prior to estimating the structural model, we can calculate the frequency at which different inventory states get visited. If a household makes a purchase that would lead them to an inadmissable state, we assume that the particular purchase doesn't change the inventory state (the bottle purchased is immediately disposed of). The rule we use is to treat the top 95% of visited states as admissable, which reduces the number of bottle states we need to track down to 23. In practice, individuals typically only purchase one or two different package sizes, and so inventory states which contain many different sizes do not get visited very often, if at all. We also include a parameter which penalizes an individual who makes a choice that leads to visiting an inadmissable state. This parameter helps us to fit choice probabilities, as without we would overpredict the probability of purchase for an individual who tries to visit such a state, since they receive utility from the error draw, but are not penalized with higher storage costs. After these restrictions, the number of states we have to track per household is 39,100. This is still four times as many as our preferred specification, where we track 10,000 states per household.

One additional note on the state space of the extended model is that we include 17 continuous inventory points, which correspond to possible volumes of detergent an individual could have for a given size of an open bottle (0 to 200 ounces). As with the zero storage cost model, we split these 17 points into two evenly-spaced grids. The first 6 points correspond to the household having 0 to 5 loads left in a bottle (given our benchmark that 100 ounces is 32 loads), while the next 11 points correspond to having 10 to 60 loads left and are evenly spaced by 5 loads (we treat the 200 ounce bottle as actually having 187.5 ounces, or 60 loads, so that the size of the bottle corresponds to the last point on the grid). We need less continuous inventory points for this model specification than the one with zero storage costs, since levels of inventory higher than 200 ounces are captured by the discrete bottle composition state, and only finitely many such configurations can occur.

Estimates of the model with increasing storage costs are shown in Table 11. The parameters SC Linear and SC Quadratic measure the effect of increasing inventory by 100 ounces on utility. The last parameter, Inadmissable State Penalty, is the disutility incurred if a consumer makes a purchase that would lead to an inadmissable state.

E Steps for implementing the IJC algorithm

Before explain the estimation details, we introduce some additional notation. Denote the vector of population-varying parameters drawn in step 1 as $\boldsymbol{\theta}_{i1}$, and the population-fixed parameters in step 2 as $\boldsymbol{\theta}_2$. We assume that the individual- specific parameters are derived from a normal distribution with mean $\boldsymbol{b}'\boldsymbol{Z}_i$ and variance $\boldsymbol{W}$, where $\boldsymbol{Z}_i$ is a matrix of demographic characteristics for household i. In the model presented in the main body of the paper, $\boldsymbol{Z}_i$ is a column vector of 1. We include additional demographics as a robustness check in the model presented in Online Appendix J. Since some of the parameters must be bounded (such as the discount factor or price coefficient) we assume that they are transformations of underlying normal parameters. We assume that the price coefficient, the stockout cost are lognormal. The transformation applied to produce the discount factor is $\exp(x)/(1+\exp(x))$, where x is normal. The inventory bound transformation is $M * \exp(x)/(1 + exp(x))$, where the maximum inventory bound M is set to be equivalent to holding 24 of the largest package size of detergent (in terms of volume, this is 4800 liquid ounces of detergent). We will denote the untransformed parameters as $\tilde{\boldsymbol{\theta}}_{i1}$, and the transformed parameters

Table 11: Dynamic Parameter Estimates: Increasing Storage Costs

Parameter	1st Tertile	Median	Mean	2nd Tertile
Price Coefficient	-0.2455	-0.1873	-0.4152	-0.1435
	[-0.2608, -0.2297]	[-0.2003, -0.1755]	[-0.5432, -0.3522]	[-0.1547, -0.1325]
Stockout Cost	0.0381	0.0726	0.2362	0.132
	[0.0235, 0.0512]	[0.0505, 0.0934]	[0.1964, 0.2735]	[0.1037, 0.1606]
Discount Factor	0	0.9613	0.5515	1
	[0, 1e-04]	[0.4529, 1]	[0.499, 0.5986]	[1, 1]
Fixed Cost of Purchase	-	-	-1.7455	-
			[-1.8525, -1.6403]	
SC Linear	0	0	0.0243	1e-04
	[0, 0]	[0, 1e-04]	[0.0079, 0.0466]	[0, 6e-04]
SC Quadratic	3e-04	0.0017	0.0776	0.0095
	[1e-04, 6e-04]	[8e-04, 0.0033]	[0.0604, 0.0949]	[0.0051, 0.017]
Inadmissable State Penalty	-	-	-1.589	-
			[-1.8008, -1.3389]	
Log-likelihood	-20840.518			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. Only the mean is shown for parameters that are fixed across the population. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Square brackets show 95% confidence intervals.

as $\boldsymbol{\theta}_{i1} = T(\tilde{\boldsymbol{\theta}}_{i1})$. Note that we assume that $\tilde{\boldsymbol{\theta}}_{i1} \sim N(\boldsymbol{b}'\boldsymbol{Z}_i, \boldsymbol{W})$.

E.1 Steps 1 to 4: Drawing the model parameters

We use the random walk Metropolis-Hastings Algorithm to implement Step 1 of the Gibbs sampler, and draw the individual specific parameters on a household-by-household basis. To that end we describe how we draw an individual $\boldsymbol{\theta}_{i1}$. Suppose that we are at step g of the Gibbs sampler. First, conditional on the last step's draw of $\tilde{\boldsymbol{\theta}}_{i1}$, which we call $\tilde{\boldsymbol{\theta}}_{i1}^0$, we draw a candidate $\tilde{\boldsymbol{\theta}}_{i1}^1$ from $N(\tilde{\boldsymbol{\theta}}_{i1}^0, \rho_1 \boldsymbol{W}_{g-1})$, where $\boldsymbol{W}_{g-1}$ is last iteration's estimate of the variance matrix. Our new utility parameters will be $\boldsymbol{\theta}_{i1}^1 = T(\tilde{\boldsymbol{\theta}}_{i1}^1)$. We then compute the joint likelihood of brand and size purchase at the old draw and the candidate draw. To implement this we first need an estimate of each consumer's value function. As we describe further in Section E.2, we compute this estimate by averaging over past value functions, using the a kernel-weighted average where the weights depend on how close the current draw is to past draws. The choice probability can be written as the probability of the observed brand choice (k_{it}) given package size choice (x_{it}) and number of bottles (j_{it}), multiplied by the probability of the observed size choice. For a given individual the probability of a particular brand choice given their choice of size is

$$Pr(k_{it}|x_{it}, p_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2) = \frac{\exp\left(\xi_{x_{it},k_{it}} - J\alpha p_{i,x_{it},k_{it},t}\right)}{\sum_{l \in C_2(x_{it})} \exp\left(\xi_{x_{it},l} - J\alpha p_{i,x_{it},l,t}\right)}.^{27} \tag{35}$$

The probability of a particular size choice can be written independently from the brand choice as

$$Pr(x_{it}, j_{it}|\boldsymbol{\Omega}_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2) = \frac{\exp\left(\frac{j_{it}}{J}\Omega_{it}(x_{it}) + \tilde{u}(I_{it}, j_{it}, x_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2) + \beta \hat{EV}_i(I_{i,t+1}, \boldsymbol{\Omega}_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2))\right)}{\sum_{(j,x)\in C} \exp\left(\frac{j_{it}}{J}\Omega_{it}(x, j) + \tilde{u}(I_{it}, j, x; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2) + \beta \hat{EV}_i(I_{i,t+1}, \boldsymbol{\Omega}_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2)\right)}, \tag{36}$$

where $\hat{EV}(I_{i,t+1}, \boldsymbol{\Omega}_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2)$ is the estimated expected value function, and

$$\tilde{u}(I_{it}, j_{it}, x_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2) = -\nu_i \frac{c_i - (I_{it} + x_{it} j_{it})}{c_i} \mathbf{1}\{I_{it} < c_i\}.^{28}$$

Note that to compute this probability we need to compute the inclusive values $\boldsymbol{\Omega}_{it}$, which themselves are functions of $\boldsymbol{\theta}_{i1}$ and $\boldsymbol{\theta}_2$ parameter draws. To construct the estimated value function we will also need to compute the transition process for the inclusive values. We discuss how the inclusive values and their transition process are computed in Section E.3.

The likelihood used for the Metropolis-Hastings accept-reject step will be

$$L_i(\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2) = \prod_{t=1}^{T_i} Pr(k_{it}|x_{it}, j_{it}, p_{it}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2) Pr(x_{it}, j_{it}|\boldsymbol{\Omega}_{it}(1); \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2).$$

The candidate draw will be accepted with probability

$$\frac{L(\boldsymbol{\theta}_{i1}^1, \boldsymbol{\theta}_2)}{L(\boldsymbol{\theta}_{i1}^0, \boldsymbol{\theta}_2)} \frac{k(\tilde{\boldsymbol{\theta}}_{i1}^0)}{k(\tilde{\boldsymbol{\theta}}_{i1}^1)},$$

where k denotes the prior density on $\boldsymbol{\theta}_{i1}$. Under our assumption of normality of the parameters this prior is simply the multivariate normal with mean $\boldsymbol{b}_{g-1}$ and variance $\boldsymbol{W}_{g-1}$.

After drawing the population-varying parameters we draw the mean ($\boldsymbol{b}$) and variance ($\boldsymbol{W}$) parameters that generate them (Steps 2 and 3). Conditional on the $\boldsymbol{\theta}_{i1}$ draws and the demographics

[27] Note that the number of bottles purchased, j_{it}, drops out of this choice probability due to the distributional assumptions made on the error term.

[28] The stockout cost enters utility slightly differently in the empirical model, to capture the idea that if a small amount of detergent is left in the bottle the consumer may use a bit of it to do laundry.

$\boldsymbol{Z}_i$, the $\boldsymbol{b}$ parameters are drawn using standard Bayesian regression. We put a relatively diffuse prior on the $\boldsymbol{b}$, using a normal distribution with mean zero and variance matrix of $1000\boldsymbol{I}$.[29]. We also assume a relatively diffuse prior on $\boldsymbol{W}$. If the dimensionality of $\overline{\boldsymbol{\theta}}_{i1,g}$ is K, then the prior variance matrix is set to $\boldsymbol{I} * (K+3) * 0.01$.[30] Given this prior, a posterior draw on the variance matrix can be computed given $\boldsymbol{b}$, $\overline{\boldsymbol{\theta}}_{i1,g}$ and $\boldsymbol{Z}_i$ from an inverse Wishart distribution.[31]

The fourth step is to draw the population-fixed parameters $\boldsymbol{\theta}_2$. This step proceeds in largely the same way as the first step. A candidate draw $\boldsymbol{\theta}_2^1$ is taken from $N(\boldsymbol{\theta}_2^0, \rho_2 \boldsymbol{W}_2)$, and is accepted with probability

$$\frac{\prod_{i=1}^{I} L_i(\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2^1)}{\prod_{i=1}^{I} L_i(\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2^0)} \frac{k(\boldsymbol{\theta}_2^0)}{k(\boldsymbol{\theta}_2^1)}.$$

We set the prior on $\boldsymbol{\theta}_2$ to be noninformative.

E.2 Step 5: Updating the value function

After a new vector of parameters are drawn, the value functions are updated at the current parameter draw. There are two steps necessary in updating the value function. First, we construct an estimate of the value function at the current parameter draw. Second, we perform a single update to the value function.

We first describe how the estimated value function is constructed. The estimated value function is constructed by integrating over the transition density of inclusive values, and by averaging over past value functions. Importantly, when we average over past value functions, we put more weight on value functions which were computed at parameter draws close to the current draw. For each individual in the data, we store the value function on a grid of 100 inventory points and 100 random price draws (meaning we update the value function for each consumer on a grid of 10,000 points). Index the inventory grid points using s_1 and the price grid points using s_2. The random prices are drawn from the empirical distribution of prices, which we denote as $h(\cdot)$. Since we assume that

[29] The dimension of $\boldsymbol{I}$ corresponds to $vec(\boldsymbol{b})$

[30] Our choice of prior variance matrix was informed by artificial data experiments, where we generated data from the myopic inventory model with heterogeneous coefficients, and recovered the parameter distributions. We found that when prior variances were too wide the estimator had some difficulty recovering small variance parameters, but scaling down the prior by a factor of about 0.01 mitigated this problem, while still allowing us to recover larger variance parameters.

[31] For more details on the process using to generate the hyperparameters we refer the readers to Rossi, Allenby, and McCulloch (2005). Our code for drawing these parameters is heavily based on the C++ code provided with the book.

the value functions are functions of inclusive values, rather than prices, the first thing we do is to compute the inclusive value at each price grid draw which arises at the current parameter draw. We denote this inclusive value as $\mathbf{\Omega}_{s_2}$.[32] Once the inclusive values are computed, we compute a set of importance weights, $w_{n_s}(\mathbf{\Omega}_{s_2})$, which are used when integrating out the transition probabilities in the value function:

$$w_{n_s}(\mathbf{\Omega}_{s_2}) = \frac{F(\mathbf{\Omega}_{n_s}|\mathbf{\Omega}_{s_2};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2)}{h(\mathbf{\Omega}_{n_s})},$$

where $F(\mathbf{\Omega}_t|\mathbf{\Omega}_{t-1};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2)$ is the transition density of the inclusive values at the current parameter draw.

The second element is to average over past value functions. For the past $g = 1, ..., G$ Gibbs draws we have $\hat{V}_i$ at each of the inventory states s_1 and the past importance draws on the inclusive values s_2. Denote the saved value functions and parameter draws respectively as $\hat{V}_i^g, \boldsymbol{\theta}_{i1}^g, \boldsymbol{\theta}_2^g, \quad g = 1, ..., G$. We compute kernel weights $\phi_g(\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2) = \phi([\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2] - [\boldsymbol{\theta}_{i1}^g,\boldsymbol{\theta}_2^g])$ where $\phi(\cdot)$ is a multivariate normal kernel function. Our value function estimate is then

$$\tilde{V}_i(I_{s_1},\mathbf{\Omega}_{s_2};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2) = \sum_{g=1}^{G}\sum_{n_s=1}^{N_s} \frac{\hat{V}^g(I_{s_1},\mathbf{\Omega}_{n_s};\boldsymbol{\theta}_{i1}^g,\boldsymbol{\theta}_2^g)\phi_g(\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2)w_{n_s}(\mathbf{\Omega}_{s_2})}{\sum_{g=1}^{G}\sum_{n_s=1}^{N_s}\phi_g(\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2)w_{n_s}(\mathbf{\Omega}_{s_2})}.$$

When we update the value function, we may need to compute the expected value function at inventory points that are not on the grid. To do this we interpolate the $\tilde{V}_i(I_{s_1},\mathbf{\Omega}_{s_2};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2)$ over inventory states using linear interpolation. Then the expected value function estimate at any inventory point I is

$$\hat{EV}_i(I,\mathbf{\Omega}_{s_2};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2) = \tilde{V}_i(\underline{I}'_{s_1},\mathbf{\Omega}_{s_2};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2) + \frac{I-\underline{I}'_{s_1}}{\overline{I}'_{s_1}-\underline{I}'_{s_1}}\tilde{V}_i(\overline{I}'_{s_1},\mathbf{\Omega}_{s_2};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2),$$

where $\underline{I}'_{s_1}$ is the largest inventory grid point that is smaller than I and $\overline{I}'_{s_1}$ is the smallest grid point that is larger than I. We will use $\hat{EV}_i(I,\mathbf{\Omega}_{s_2};\boldsymbol{\theta}_{i1},\boldsymbol{\theta}_2)$ when updating the value function, as we describe in the next paragraph. Before moving on, we also note that when we compute choice probabilities in Section E.1 we also need to compute the expected value function; for this we use a similar procedure to the above (the main difference is that the value function approximation is evaluated at an inclusive value derived from an observed price, rather than a particular inclusive value grid point).

[32] An alternative approach we experimented with was to propose an importance distribution for inclusive values and to draw inclusive values from that. We found that this approach would sometimes lead to numerical errors when computing transition probabilities, if the current parameter draw was very far from all of the drawn inclusive values. (maybe move this down)

Denote this value function estimate as $\hat{EV}_i(I, \boldsymbol{\Omega}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2)$. We will index inventory grid points (which are fixed across Gibbs iterations) with s_1 and inclusive value grid points (which are random) with s_2. Then at a particular grid point s_1, s_2, with state variables $I_{s_1}, \boldsymbol{\Omega}_{s_2}$, the updated value function is

$$\hat{V}_i(I_{s_1}, \boldsymbol{\Omega}_{s_2}; \boldsymbol{\theta}_{i1}; \boldsymbol{\theta}_2) = \sum_{(j,x)\in C} \log\left(\exp\left(\hat{\Omega}_{s_2}(x,j) + \tilde{u}(I_{s_1}, j, x; \boldsymbol{\theta}_{i1}; \boldsymbol{\theta}_2) + \beta \hat{EV}_i(I'_{s_1}, \boldsymbol{\Omega}_{s_1}; \boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_2)\right)\right). \tag{37}$$

E.3 Inclusive value transition process

When we take a new draw on the parameters $\boldsymbol{\theta}_{i1}$ and $\boldsymbol{\theta}_2$ we need to compute new inclusive values, as well as to estimate their transition processes, and the functions for approximating the inclusive values for $j > 1$. Our approach builds on Hendel and Nevo (2006a) in two ways. First, we estimate the except we estimate the inclusive value transition process at the individual rather than population level. Second, rather than using a linear AR1 specification for the inclusive value transition process, we use a more flexible spline process which we believe should fit the transition process of prices better. In particular, if a particularly low price is observed the transition process is probably different than a high price.

The spline basis functions we use are the B-spline basis functions proposed in (**?**). The procedure we use works as follows: first, for every individual i, week t and package size k we compute inclusive values, which we denote as $\Omega_{k,it}$. Then for each individual and package size we calculate the maximum and minimum values of the inclusive values, and decide on how many knot points will be included. We choose to use $n = 1$ knots, which means that the transition process will depend on whether the inclusive value is below the midpoint, or above it. This approach seems reasonable since the price process of laundry detergents seems to follow a hi-lo process, with price being at a regular retail price for some time and periodically dipping to a very low value. Given n knot points, the procedure produces $n+2$ basis functions. We denote the basis function l for individual i, and package k as $\phi_{k,l,i}(\cdot)$. Then the regression equation that

$$\Omega_{k,it} = \sum_{j=1}^{K} \sum_{l=1}^{n+2} \kappa_{k,l,i} \phi_{k,l,i}(\Omega_{j,it-1}) + \varepsilon_{it},$$

where for each size $j \neq k$ we must normalize one $\kappa_{k,l,i}$ to 0. This normalization is necessary because the sum of the basis functions for a given size equals $n+2$ by construction. We choose to normalize the coefficient for basis function $n+2$. We note that we do not need to normalize any parameter for size k, since we do not include a constant in the regression.

When implementing this procedure, we found that for a few individuals and package sizes the basis functions would become close to collinear if that individual observed very little price variation for that size, or if the likelihood was being evaluated at a particularly extreme draw of the parameters, which would lead to numerical instability in the code which would solve for the regression parameters. To avoid this problem, after constructing the matrix of independent variables during estimation we compute the condition number of the $X'X$ matrix, and if the $X'X$ matrix has a condition number above 1×10^{12} we run an AR1 regression instead of the spline regression above.

E.4 Setup of the Gibbs Sampler

In this section we describe some details of the setup of the Gibbs sampler. The computer code we use is written in R and C++ and designed to take advantage of parallel processing in the value function averaging and updating. Our code for Bayesian estimation makes use of routines from Rossi, Allenby, and McCulloch (2005) for summarizing the model output as well as for drawing the heirarchical parameters. For the Metropolis-Hastings steps in steps 1 and 4 of the Gibbs sampler we need to set the parameters ρ_1 and ρ_2, which control the variance of the random walk process for the population-varying and population-fixed parameters, respectively. Each of these parameters are tuned so that the acceptance rate over the course of the sampler is about 30%. We tune the parameter ρ_1 every iteration: if the fraction of household level parameters that are accepted is above 30%, we increase ρ_1 by 10%; otherwise we decrease it by 10%. For ρ_2, we adjust the parameter every 25 iterations: if the number of acceptances for the past 25 iterations is above 30%, then the ρ_2 parameter is decreased by 25%; otherwise it is increasd by 25%. The ρ parameters move some initially but settle down after about 500 iterations.

To compute the $\boldsymbol{W}_2$ matrix, we estimate the dynamic stockpiling model assuming no unobserved parameter heterogeneity using MCMC (there is still heterogeneity in consumer price expectations and consumption rates). Then, we compute the inverse information matrix of the likelihood function evaluated at these initial estimates. We set $\boldsymbol{W}_2$ to be the submatrix of the inverse information matrix corresponding to the parameters which are set to be fixed across the population. We found this procedure for setting $\boldsymbol{W}_2$ worked well in artificial data experiments.

For the value function approximation we choose $G = 10$, and we use a diagonal bandwidth matrix with bandwidth parameter set to $(4/(3G))^{0.2}$. We found these values worked well in artificial data experiments where we tested out the sampler. As we discussed above we evaluate the value function on 100 grid points. Grid points for inventory are chosen between 0 and the maximum inventory value of 4800 ounces, with the first 20 points of the points being clustered equally between

0 and 70 ounces (around the size of a smaller package of detergent) and the rest between 700 and 4800. We choose more points near zero since the incentive to purchase in advance of running out becomes more important for low inventory levels, and we want to make sure we capture that behavior well.[33] We run the Gibbs sampler for 20,000 iterations. The draws appear to converge at about 4,000 iterations, so we drop the first 4,000 draws to reduce burn-in.

F Model Fits

In this section we compare how the main empirical model specification, where discount factors are free, fits some selected moments of the data relative to the model specification that fixes the discount factor to the rational expectations benchmark. We emphasize that overall, the model with free discount factors provides a significantly better fit to overall purchases, which is a result of it providing a higher log-likelihood and lower Deviance Information Criterion. In Table 12, we compare summary statistics related to purchase timing, such as the overall purchase probability and the number of weeks between purchases. The first row shows the overall purchase probability - the first column is estimated from the data and the second and third show the simulated probabilities from the model. The models underestimate the purchase probability by a small amount, but the model with flexible discount factor fits the data a bit better than the rational expectations model. The next rows show the summary statistics of interpurchase timing, measured in weeks. Overall, the structural model seems to mimic the percentiles of the empirical distribution of interpurchase timing, but systematically overpredicts the time between purchases by about 1 to 2 weeks. Generally, the model with the free discount factor is a little bit closer to the data. In Table 13, we show the actual and predicted purchase shares for all products available in the data. The predicted shares are close to the truth, typically within 1% or less.

We examine how imputed inventory varies with actual and predicted purchase probabilities in Table 14. The first column shows the empirical purchase probability given a particular range of imputed inventory. We present imputed inventory in terms of hundreds of ounces, or, equivalently, the number of one hundred ounce bottles held. As one would expect, the empirical probabilities drop as inventory rises, consistent with stockpiling behavior. In the second and third columns, we show the average of the predicted purchase probabilities for the same observations use to construct the estimates in the data column. The purchase probabilities are constructed for each individual using the average of their posterior draws. The main model specification replicates this trend,

[33] Equivalently, most of the nonlinearity in the value function occurs for low inventory levels, so having more interpolation points in that region of the state space ensures our approximation to the value function is good.

Table 12: Fit: Summary Statistics Related to Purchase Timing

Moment	Data	Main Model	Rational Expectations
Overall Purchase Prob.	10.17	8.74	8.67
Interpurchase 25th Percentile	4	4	4.01
Interpurchase 50th Percentile	7	8.42	8.68
Interpurchase Mean	9.89	11.59	11.68
Interpurchase 75th Percentile	13	15.4	15.6
Interpurchase Standard Dev.	8.06	10.97	10.93

Notes: In the data column, purchase prob measures the overall empirical probability of a purchase in the estimation data. The interpurchase rows refer to the percentiles, mean and standard deviation of the number of weeks between purchases in the empirical data. In the second and third columns, we simulate the corresponding moments out of each model. The simulation is done for every MCMC draw and simulated choices are saved for every 10th draw to save hard drive space. The predicted summary statistics are computed for choices that are saved and averaged across draws.

although it underpredicts the purchase probability at low levels of inventory somewhat. The model specification that corresponds to the rational expectations benchmark displays a somewhat different trend. It predicts higher purchase probabilities at low levels of inventory, which is more consistent with the data, but also overpredicts purchase at higher levels of inventory, which is inconsistent with the data. What is likely happening here is that if one assumes rational expectations, consumers will feel stronger incentives to hold higher levels of inventory and make purchases, even if they are already carrying a lot of a product.

Our counterfactual exercise suggests that the rational expectations model overpredicts how individuals respond to drops in price, because more forward-looking individuals have stronger incentives to time their purchases to occur during discounts. We present evidence in support of this intuition in a series of regressions shown in Table 15. In these regressions, we regress a measure of the inclusive value for the most popular 100 ounce size on a purchase dummy variable (first column), and the predicted likelihood of purchase on this inclusive value. The inclusive value we use is constructed from the estimates of a homogeneous parameters version of the model, which we use as starting points for both our estimation specifications (we construct the inclusive value in this way because we wanted to use an estimate that was independent of either of the model specifications). The purchase likelihood is the predicted likelihood of a purchase at each observation in the data, and is again constructed for each individual using the average of the individual's posterior draws. As expected, a higher inclusive value is correlated with an increased likelihood of purchase. The

Table 13: Fit: Product-Level Shares

Moment	Data	Main Model	Rational Expectations
Tide, 100 OZ	19.09	16.96	17.65
Tide, 200 OZ	5.19	6.21	5.89
Tide, 50 OZ	2.25	3.09	3.23
Tide, 80 OZ	0.08	0.14	0.17
Xtra, 128 OZ	9.44	9.34	9.26
Xtra, 200 OZ	0.8	1.3	1.27
Purex, 100 OZ	7.87	7.53	7.45
Purex, 128 OZ	0.03	0.1	0.1
Purex, 200 OZ	0.82	0.79	0.7
Purex, 50 OZ	0.11	0.25	0.2
All, 100 OZ	8.5	6.45	6.29
All, 200 OZ	0.47	0.68	0.63
All, 50 OZ	0.22	0.99	0.92
All, 80 OZ	0.03	0.08	0.07
Arm & Hammer, 100 OZ	7.24	6.64	6.61
Arm & Hammer, 200 OZ	0.69	0.65	0.61
Era, 100 OZ	5.32	4.67	4.71
Era, 200 OZ	0.71	0.27	0.23
Era, 50 OZ	0.14	0.34	0.34
Dynamo, 100 OZ	6.45	6.49	6.54
Dynamo, 200 OZ	0.14	1.16	1.16
Wisk, 100 OZ	11.63	8.14	8.01
Wisk, 200 OZ	0.03	0.45	0.44
Wisk, 80 OZ	0.14	0.6	0.6
Private Label, 100 OZ	1.51	2.09	2.19
Private Label, 128 OZ	2.77	2.81	2.84
Private Label, 50 OZ	0.03	0.11	0.1
Cheer, 80 OZ	2.41	3.31	3.44
Fab, 100 OZ	0.52	0.95	0.9
Fab, 50 OZ	0.6	0.2	0.19
Yes, 100 OZ	1.29	1.25	1.3
Ajax Fresh, 128 OZ	0.47	0.69	0.67
Gain, 100 OZ	0.63	0.69	0.71
Gain, 200 OZ	0	0.01	0.01
Ajax, 128 OZ	0.47	1.7	1.7
Trend, 128 OZ	0.47	0.96	0.99
Sun, 100 OZ	0.47	0.48	0.49
Solo, 100 OZ	0.63	0.83	0.8
Ivory Snow, 50 OZ	0.36	0.59	0.6

Notes: In the data column, we show the overall purchase share of each product available, given a purchase occurs. In the second and third columns, we simulate the corresponding moments out of each model. The simulation is done for every MCMC draw and simulated choices are saved for every 10th draw to save hard drive space. The predicted summary statistics are computed for choices that are saved and averaged across draws.

Table 14: Fit: Purchase Probabilities for Different Levels of Imputed Inventory

Number of 100 ounce Bottle Equivalent	Data	Main Model	Rational Expectations
0-1	16.41	11.49	13.15
1-2	10.59	7.75	8.58
2-3	8.31	7.63	8.38
3-4	9.32	8.15	9.33
4-5	8.54	8.65	10.22
5-6	8.46	8.91	10.17
6+	8.42	8.91	10.03

Notes: In the data column, purchase prob measures the overall empirical probability of a purchase in the estimation data for a given level of imputed inventory (in hundreds of ounces). In the second and third columns, we show the average predicted purchase likelihoods, at the average of the individual-level draws, for each level of imputed inventory in the data.

Table 15: Fit: Regression of Purchase Indicator/Likelihood on Inclusive Value

Coefficient	Data	Main Model	Rational Expectations
Intercept	0.2098	0.1953	0.2428
	(0.01387)	(0.00211)	(0.00277)
Inclusive Value	0.0185	0.0183	0.0245
	(0.00236)	(0.00036)	(0.00047)

Notes: In the data column, we present a regression of a purchase dummy variable on the estimated inclusive value . In the second and third columns, we show the predicted purchase likelihoods, at the average of the individual-level draws, for each level of imputed inventory in the data.

main model specification with the free discount factor has a correlation that is very similar in magnitude. However, the correlation between the inclusive value and the purchase likelihood in the rational expectations model is much larger. The fact that this model overstates price sensitivity is consistent with the overprediction of quantity increase due to increased promotion depth in our counterfactual exercise.

G Artificial Data Experiment: Empirical Model

In this section we will provide additional support in our data that we can identify the model parameters, including the discount factor, and heterogeneity in the discount factor. We will do this by simulating consumer choices from the model, and running the MCMC sampler to recover the parameters. Our procedure for generating the parameters for our simulation is to compute the average mean and variance matrix over saved draws of the individual-specific parameters, without any transformations applied, and then to draw individual-specific parameters from a normal distribution with the estimated mean and variance. We then apply parameter transformations (i.e., on the discount factor or price coefficient), to produce the individual-specific parameters. We then solve for every individual's value function, and simulate individual choices. Choices are simulated for both the pre-estimation sample (assuming zero inventory at the beginning of the individual's time series), and the estimation sample, for all individuals. The MCMC sampler is then run on the simulated data in exactly the same way as it is run on the estimation data.

The estimation results for the dynamic parameters are shown in Table 16. The first column shows the average of the population distribution of individual-level parameters, along with confidence bounds in brackets, while the column beside it shows the true mean of the individual-level parameters. The estimated model parameters are very close to the truth, and the truth is within the confidence bounds of the estimates. The next two columns show the estimated population standard deviations, along with the actual standard deviations. Again, the true values are within the confidence bounds of the estimates. We find similar results for the estimates of the product-specific parameters, and their heterogeneity, shown in Online Appendix Tables 17 and 18.

Table 16: Dynamic Parameter Estimates: Artificial Data Experiment

Parameter	Estimated Mean	True Mean	Estimated Standard Deviation	True Standard Deviation
Price Coefficient	-0.27	-0.27	0.11	0.1
	[-0.28, -0.26]		[0.1, 0.14]	
Stockout Cost	0.42	0.42	0.19	0.16
	[0.36, 0.48]		[0.14, 0.27]	
Discount Factor	0.85	0.88	0.11	0.16
	[0.81, 0.91]		[0.05, 0.16]	
Fixed Cost of Purchase	-1.82	-1.83	-	-
	[-1.91, -1.74]			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters (columns 1 and 3) compared to the true moments of the individual-level parameters used to construct the simulated data. Only the mean is shown for parameters that are fixed across the population. Square brackets show 95% confidence intervals.

Table 17: Product Parameter Estimates: Artificial Data Experiment (1)

Parameter	Estimated Mean	True Mean	Estimated Standard Deviation	True Standard Deviation
Tide, 200 OZ	-2.05	-2.24	-	-
	[-2.61, -1.48]			
Xtra	-13.4	-13.48	4.58	4.52
	[-14.11, -12.73]		[4, 5.16]	
Xtra, 200 OZ	-12.52	-13.53	-	-
	[-13.11, -11.86]			
Purex	-5.55	-5.68	2.69	2.57
	[-5.88, -5.26]		[2.38, 3.03]	
Purex, 128 OZ	-19.79	-18.27	-	-
	[-22.97, -17.45]			
All	-4.65	-4.86	3.19	2.99
	[-5.03, -4.35]		[2.85, 3.61]	
Arm & Hammer	-5.24	-5.21	2.78	2.49
	[-5.54, -4.97]		[2.47, 3.14]	
Era	-6.95	-6.88	3.1	2.89
	[-7.57, -6.39]		[2.63, 3.64]	
Dynamo	-6.78	-6.54	3.67	3.38
	[-7.29, -6.29]		[3.23, 4.18]	
Wisk	-4.14	-4.22	3.24	3.05
	[-4.69, -3.67]		[2.83, 3.86]	
Private Label	-12.01	-11.49	4.19	3.88
	[-12.82, -11.1]		[3.42, 4.81]	
Cheer	-13.37	-13.59	-	-
	[-14.26, -12.54]			
Fab	-8.95	-9.04	3.68	4.63
	[-9.58, -8.26]		[3.13, 4.43]	

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Only the mean is shown for parameters that are fixed across the population. Square brackets show 95% confidence intervals.

Table 18: Product Parameter Estimates: Artificial Data Experiment (2)

Parameter	Estimated Mean	True Mean	Estimated Standard Deviation	True Standard Deviation
Yes	-8.2	-7.23	3.85	2.91
	[-8.94, -7.46]		[3.17, 4.56]	
Ajax Fresh	-16.99	-16.4	1.88	4.3
	[-18.57, -15.71]		[1.2, 2.44]	
Gain	-5.47	-5.42	-	-
	[-5.87, -5.06]			
Ajax	-13.93	-14.02	-	-
	[-14.62, -13.16]			
Trend	-13.77	-14.03	-	-
	[-14.52, -12.96]			
Sun	-6.54	-6.77	-	-
	[-7.02, -6.11]			
Solo	-7.94	-7.6	-	-
	[-8.37, -7.5]			
Ivory Snow	-14.78	-14.73	-	-
	[-15.69, -13.87]			
128 OZ	-4.2	-4.05	-	-
	[-4.91, -3.46]			
200 OZ	-5.24	-5.7	-	-
	[-5.85, -4.66]			
50 OZ	-12.97	-12.86	-	-
	[-13.72, -12.23]			
80 OZ	-14.39	-15.42	-	-
	[-15.42, -13.36]			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Only the mean is shown for parameters that are fixed across the population. Square brackets show 95% confidence intervals.

H Supplementary Tables and Figures

Figure 4: Price (in dollars) of 200 OZ Tide at a Representative Store

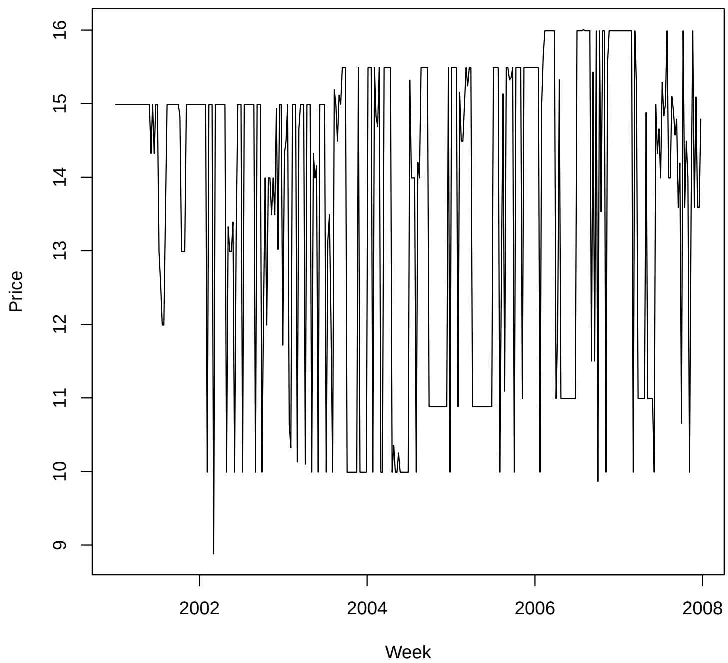

Table 19: Average Percent Price Reduction, Percent of Deals for Available Brand-Sizes

Brand	100 oz	128 oz	200 oz	50 oz	80 oz
Tide	20, 34	-	18, 33	15, 9	9, 2
Xtra	-	18, 24	18, 15	-	-
Purex	30, 33	16, 27	27, 9	30, 21	-
All	22, 33	-	15, 14	18, 10	27, 18
Arm & Hammer	26, 26	39, 14	30, 10	6, 18	-
Era	22, 21	-	23, 18	45, 2	-
Dynamo	30, 27	-	31, 17	20, 27	-
Wisk	21, 33	-	16, 16	23, 23	28, 42
Private Label	20, 22	12, 11	10, 12	33, 10	-
Cheer	16, 12	-	-	42, 1	21, 16
Fab	33, 18	-	-	33, 31	-
Yes	38, 20	-	-	3, 31	-
Ajax Fresh	-	23, 16	-	-	-
Gain	28, 19	-	28, 40	-	-
Ajax	6, 36	24, 20	-	-	-
Trend	-	-	-	-	-
Sun	-	-	-	-	-
Solo	23, 61	25, 6	4, 32	-	-
Ivory Snow	4, 64	-	-	-	-

Notes: The left number is the average percent difference between the discounted price and the non-discounted price in a store, while the right shows the percentage of store-week observations where a discount occurs. The numbers in this table are constructed from IRI's store panel data. The number of observations is 25550 store-week-product combinations.

Table 20: Brand Parameter Estimates (1)

Parameter	1st Tertile	Median	Mean	2nd Tertile
Tide, 200 OZ	-	-	-2.24	-
			[-2.8, -1.66]	
Xtra	-15.99	-13.35	-13.31	-10.86
	[-17.53, -14.75]	[-14.4, -12.42]	[-14.37, -12.4]	[-11.66, -10.07]
Xtra, 200 OZ	-	-	-13.53	-
			[-14.37, -12.7]	
Purex	-6.94	-5.35	-5.51	-3.96
	[-7.71, -6.2]	[-5.87, -4.79]	[-6.1, -4.96]	[-4.39, -3.53]
Purex, 128 OZ	-	-	-18.27	-
			[-21.28, -15.88]	
All	-6.46	-4.69	-4.87	-3.13
	[-7.34, -5.73]	[-5.28, -4.14]	[-5.47, -4.38]	[-3.6, -2.67]
Arm & Hammer	-6.48	-4.99	-5.11	-3.69
	[-7.28, -5.89]	[-5.57, -4.56]	[-5.74, -4.7]	[-4.14, -3.3]
Era	-8.51	-6.69	-6.7	-4.96
	[-9.43, -7.74]	[-7.4, -6.08]	[-7.38, -6.17]	[-5.55, -4.43]
Dynamo	-8.24	-6.18	-6.34	-4.35
	[-9.09, -7.44]	[-6.8, -5.53]	[-6.9, -5.82]	[-4.82, -3.9]
Wisk	-6.04	-4.31	-4.36	-2.7
	[-7.27, -5.16]	[-5.21, -3.68]	[-5.2, -3.75]	[-3.28, -2.22]
Private Label	-13.83	-11.42	-11.42	-9.14
	[-16.51, -12.23]	[-13.43, -10.15]	[-13.51, -10.22]	[-10.55, -8.15]
Cheer	-	-	-13.59	-
			[-14.52, -12.65]	
Fab	-11.23	-8.41	-8.85	-6.04
	[-12.86, -9.79]	[-9.65, -7.34]	[-10.09, -7.87]	[-7.02, -5.11]

Notes: The first column of the table show the average of the estimated posterior distribution of the brand parameters. Only the mean is shown for parameters that are fixed across the population. The second shows the 95% confidence bound around the mean. Brand coefficients for Tide (the most popular product) are normalized to be zero across the population.

Table 21: Brand Parameter Estimates (2)

Parameter	1st Tertile	Median	Mean	2nd Tertile
Yes	-8.8	-6.97	-7.09	-5.29
	[-9.67, -8.12]	[-7.54, -6.47]	[-7.64, -6.66]	[-5.74, -4.8]
Ajax Fresh	-18.77	-16.05	-16.2	-13.68
	[-20.43, -17.07]	[-17.1, -14.86]	[-17.27, -15]	[-14.5, -12.83]
Gain	-	-	-5.42	-
			[-5.89, -5.01]	
Ajax	-	-	-14.02	-
			[-14.93, -13.2]	
Trend	-	-	-14.03	-
			[-15.02, -13.12]	
Sun	-	-	-6.77	-
			[-7.42, -6.18]	
Solo	-	-	-7.6	-
			[-8.08, -7.11]	
Ivory Snow	-	-	-14.73	-
			[-15.82, -13.78]	
128 OZ	-	-	-4.05	-
			[-4.83, -3.15]	
200 OZ	-	-	-5.7	-
			[-6.32, -5.08]	
50 OZ	-	-	-12.86	-
			[-13.69, -11.99]	
80 OZ	-	-	-15.42	-
			[-16.7, -14.26]	

Notes: The first column of the table show the average of the estimated posterior distribution of the brand parameters. Only the mean is shown for parameters that are fixed across the population. The second shows the 95% confidence bound around the mean. Brand coefficients for Tide (the most popular product) are normalized to be zero across the population.

I Supplementary Identification Figures

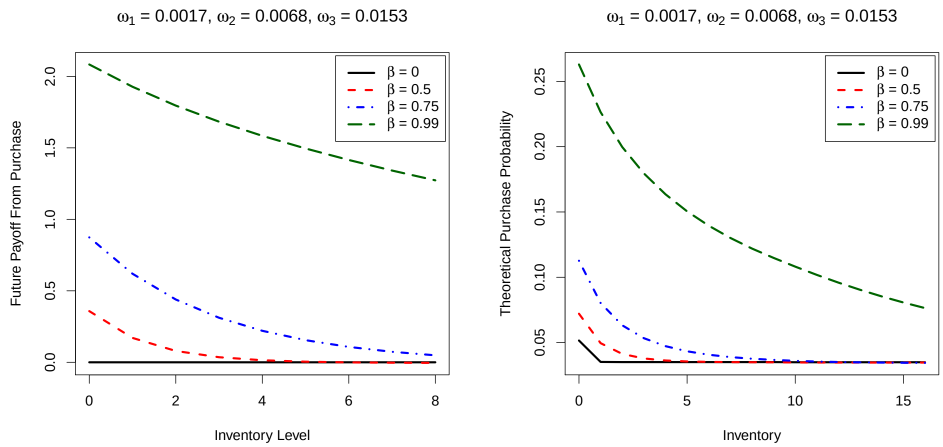

Figure 5: Expected future payoff from purchase, $\beta\,[V(I_1'(I)) - V(I_0'(I))]$ (left panel), and purchase probability (right panel), as a function of I and β, for storage costs estimated in Online Appendix J, Table 11 from the model in Online Appendix D. Parameter values $\nu = 0.4$, $M = 3$, $p = 3.31$, and logit error term.

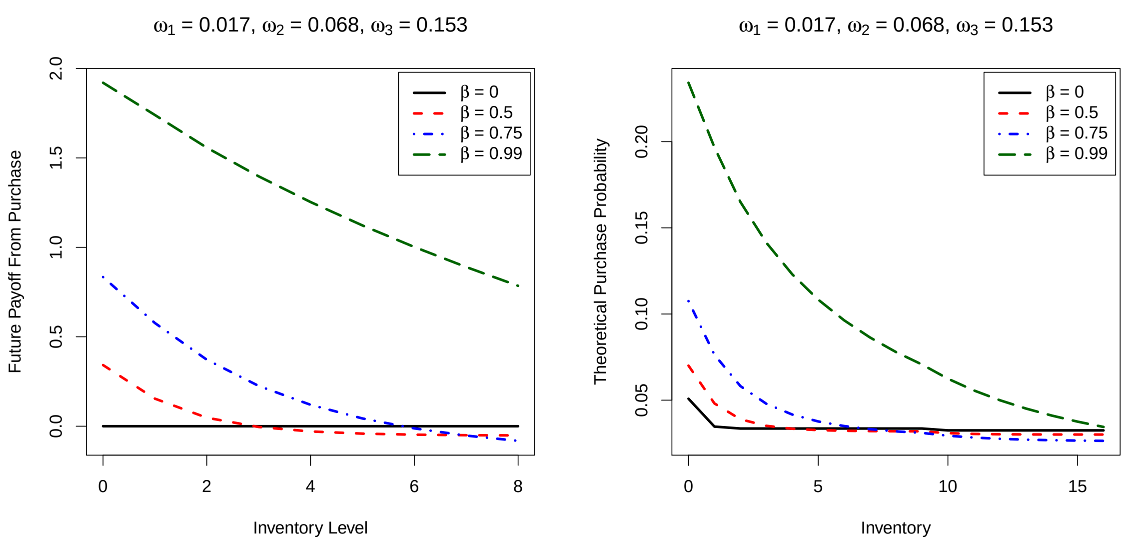

Figure 6: Expected future payoff from purchase, $\beta\,[V(I_1'(I)) - V(I_0'(I))]$ (left panel), and purchase probability (right panel), as a function of I and β, for storage costs ten times those in Figure 5. Parameter values $\nu = 0.4$, $M = 3$, $p = 3.31$, and logit error term.

To construct Figure 9, we first compute for every individual in the data their value function

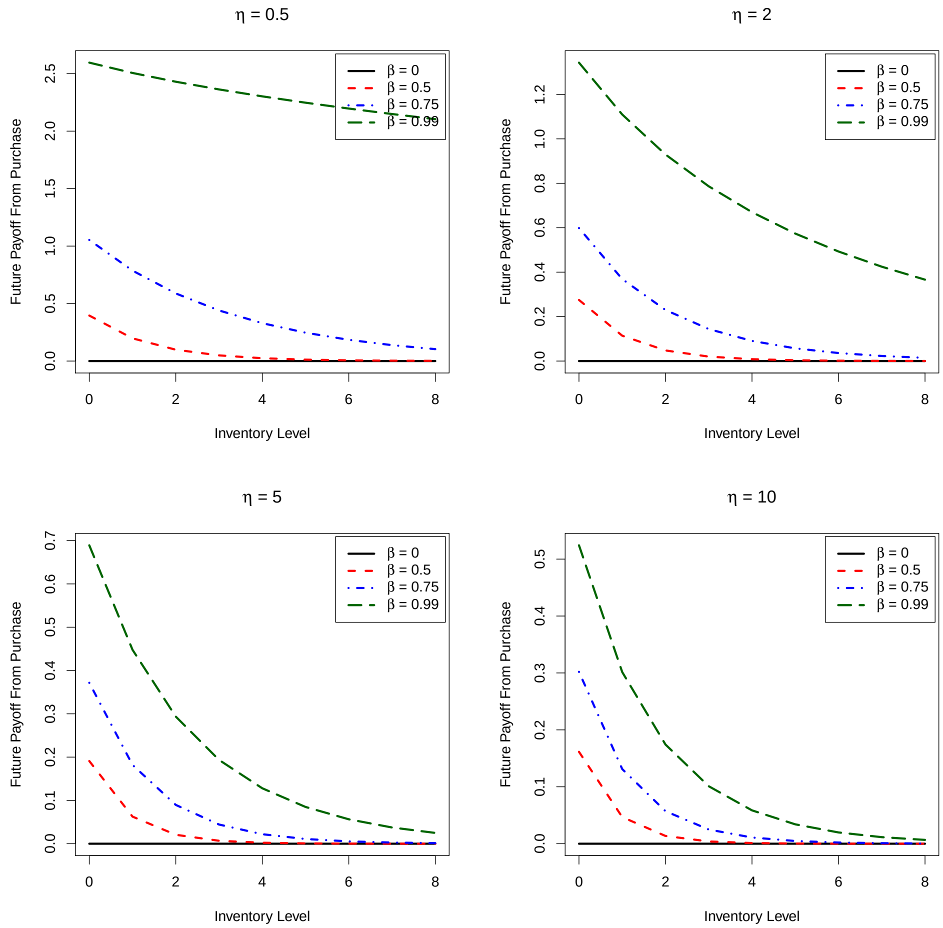

Figure 7: Expected future payoff from purchase, $\beta\,[V(I_1'(I)) - V(I_0'(I))]$, as a function of I and β, for different weights on the error term (we multiply the logit error by a parameter η). Parameter values $\nu = 0.33, \omega_1 = \omega_2 = \omega_3 = 0$, $M = 3$, $p = 1.77$, and logit error term.

and choice probabilities at the average of their individual-specific parameter draws, for each inventory and price state. Then, we compute the average expected future payoff by averaging across individuals and price states at each inventory state point. When we average across price states, we weight each point by the empirical probability of that price occurring in the household purchase data. When computing the expected future value of purchase at next week's inventory given today's inventory, we compute the value function at the expected level of inventory an individual would have tomorrow, if they purchased today. To compute this average we use the individual's probabilities of choosing each size and number of bottles given the individual's parameters and the price state.

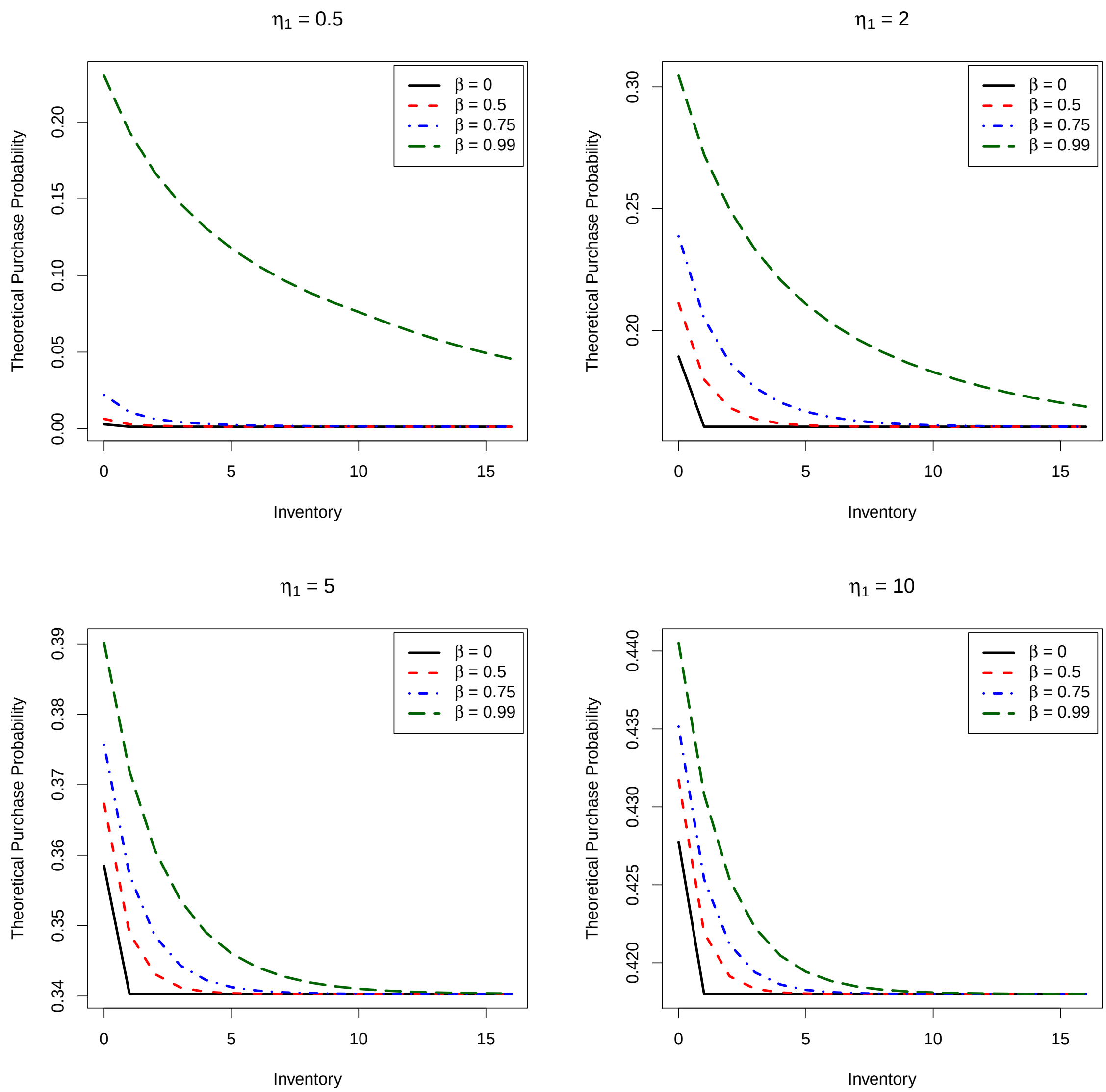

Figure 8: Theoretical probability of purchase for a given level of inventory at the beginning of a period, for different weights on the error term (we multiply the logit error by a parameter η). Parameter values $\nu = 0.33, \omega_1 = \omega_2 = \omega_3 = 0$, $M = 3$, $p = 1.77$, and logit error term.

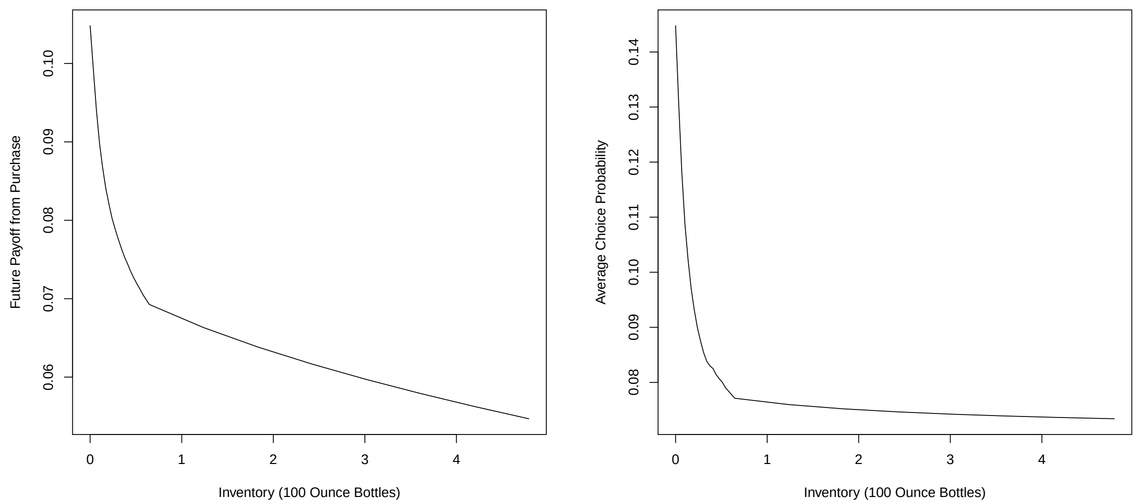

Figure 9: Average Expected Future Payoff from Purchase (Left Panel), and Average Predicted Purchase Probability (Right Panel), Computed from the Estimation Model.

J Supplementary Estimation Results

In this section we present estimation results from all of our additional specifications, as well as additional results from the main specification. Table 23 contains the dynamic parameter estimates resulting from increasing consumption rates by 25%.

We also estimated a version of the model with demographic interactions. Sample averages of these demographic variables are shown in Table 22. We include four demographic variables, all of which are coded as dummy variables. The income variable codes whether the household's income is above \$35,000 (the median in our estimation sample), whether the household head's age is about 55 years (also the median household age in the sample), whether the household head has a college degree, and whether the household has 3 or more individuals in it. The estimation sample somewhat oversamples elderly households and households with 2 individuals, relative to the U.S. population.

In Table 24, we present the overall estimated coefficients from this specification. The results are qualitatively similar to the results in our main specification in the paper, in particular, we still find that there is a significant spread in the discount factors with a sizeable share of forward-looking individuals. Table 25 shows the estimated marginal impact of each demographic variable on all the parameters which are allowed to vary across the population. In estimation, the untransformed mean vector of these parameters for household i is $\boldsymbol{b}'\boldsymbol{Z}_i$, where $\boldsymbol{Z}_i$ is a matrix of demographic characteristics for the household and $\boldsymbol{b}$ is a matrix of underlying mean parameters. The first column

Table 22: Averages of Demographic Dummy Variables

Variable	Average
HH Income $\geq$ \$35,000	0.59
HH Head Age $\geq$ 55	0.64
HH Head has college degree	0.21
HH Size 3+	0.32

Table 23: Dynamic Parameter Estimates: Consumption Rates Increased by 25%

Parameter	1st Tertile	Median	Mean	2nd Tertile
Price Coefficient	-0.3198	-0.2546	-0.468	-0.2084
	[-0.3428, -0.2982]	[-0.2702, -0.2407]	[-0.568, -0.3484]	[-0.2207, -0.1974]
Stockout Cost	0.2507	0.3703	0.6424	0.5412
	[0.2028, 0.3064]	[0.3135, 0.428]	[0.5523, 0.7323]	[0.4618, 0.6277]
Discount Factor	0.1084	0.9458	0.5975	0.9997
	[0.0078, 0.3979]	[0.8045, 0.9942]	[0.5447, 0.6526]	[0.9982, 1]
Fixed Cost of Purchase	-	-	-2.0976	-
			[-2.221, -1.9915]	
Log-likelihood	-20896.8188			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. Only the mean is shown for parameters that are fixed across the population. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Square brackets show 95% confidence intervals.

of this table shows the estimated parameter at the modal value of the demographic variables. For example, if we denote the modal demographics in matrix form as $\overline{\boldsymbol{Z}}$, then for the price coefficient we show in the first column $\exp(\boldsymbol{b}_l'\overline{\boldsymbol{Z}})$, where l is the row of $\boldsymbol{b}$ corresponding to the untransformed price coefficient. The table shows this value averaged across saved draws. Each column shows how the estimated parameter changes when the corresponding demographic variable is changed from zero to one. For example, if a household is high income its price coefficient is closer to zero by 0.009 (i.e., higher income households are more price sensitive). Most of the interactions are not statistically different from zero, which accords with some of our artificial data experiments, where we found it difficult to identify the coefficients of the demographic interactions.

Table 24: Dynamic Parameter Estimates: Including Demographic Interactions

Parameter	1st Tertile	Median	Mean	2nd Tertile
Price Coefficient	-0.2933	-0.2437	-0.2746	-0.2062
	[-0.307, -0.2802]	[-0.2542, -0.2333]	[-0.2854, -0.2637]	[-0.2152, -0.1971]
Stockout Cost	0.3568	0.4925	0.6583	0.6697
	[0.2796, 0.441]	[0.3897, 0.5937]	[0.5252, 0.7871]	[0.5363, 0.8034]
Discount Factor	0.086	0.7901	0.5634	0.9976
	[1e-04, 0.474]	[0.0829, 0.9953]	[0.441, 0.6625]	[0.979, 1]
Fixed Cost of Purchase	-	-	-1.8309	-
			[-1.8987, -1.7565]	
Log-likelihood	-19569.9431			

Notes: This table shows average moments of the posterior distribution of the population distribution of the dynamic parameters. Only the mean is shown for parameters that are fixed across the population. For example, the median columns shows the average of the population median of a given parameter, where the average is taken across MCMC draws. Square brackets show 95% confidence intervals.

Table 25: Marginal Effects of Demographic Variables

Parameter	Baseline	HH Income	HH Head Age	HH Head College	HH Size
Price Coefficient	-0.25	0.042	-0.045	0	-0.006
		[0.006, 0.077]	[-0.077, -0.01]	[-0.036, 0.036]	[-0.048, 0.032]
Stockout Cost	0.469	-0.036	-0.176	0.258	-0.181
		[-0.246, 0.158]	[-0.445, 0.139]	[-0.024, 0.557]	[-0.312, -0.001]
Discount Factor	0.563	-0.435	0.497	-0.037	0.243
		[-0.993, 0]	[0.001, 0.966]	[-0.838, 0.754]	[-0.204, 0.889]

Notes: This table shows the estimated impact of changing one of the demographic dummy variables from zero to one on a particular parameter. The respective demographic dummy variables are defined to be 1 under the following conditions: Income above $35,000; age of household head above 55; household head has a college degree; size of household is more than 2 individuals. The baseline column shows the predicted value of a parameter at the mode of the demographic distribution. The modal values are high income, older household head, no college degree, and two individuals in the household.

K Formal Propositions Related to Identification Discussion

In this section we provide formal conditions to guarantee when the future value of purchase decreases in inventory, and increases in the discount factor, as discussed in Section 4. These two properties are difficult to formally prove in situations where there is a choice-specific error that is unobserved to the researcher, because there are generally no closed form solutions for the consumer value functions. Therefore, our proof involves two steps. We augment the model from Section 3 by multiplying the choice-specific error in equation (2) by a weight $\eta \geq 0$, and first, we prove the properties are true in a simple setting where the error term has zero variance ($\eta = 0$); in this situation we can derive analytical formulas for consumer policy and value functions. Second, we demonstrate that the value function is continuous in η under some regularity conditions on the error term. Hence, as long as payoffs is bounded, it will be the case that in some neighborhood of $\eta = 0$, the two properties will still hold.

To keep the notation and the proofs analytically tractable, we will make a few more simplifying assumptions, some of which we can relax. We consider the case where all the model parameters are homogeneous across the population, and thus will drop the i subscript on everything except the state variables and the error term. We also normalize the consumption utility, γ, to 0, as we mention in Section 3. We also make five simplifying assumptions below:

Assumptions A1'-A5'

1. The consumption rate is constant across time and individuals: $c_{it} = 1$ for all i, t.
2. In a given purchase occasion, the maximum number of packages a consumer is allowed to buy is 1.
3. Prices are fixed over time at a level $p > \beta^{b-1}\omega_1 \geq 0$.
4. Purchasing at 0 inventory is preferred to running out: $\frac{1-\beta^{b-1}}{1-\beta}\omega_1 + p < \nu$.
5. The storage cost function is weakly increasing and weakly convex, and storage costs are weakly positive.

Assumption A1' is made for convenience; it can be shown that the propositions in this section will still hold under stochastic consumption needs (proofs are available upon request). Assumptions A2' and A3' are made for tractability. A2' ensures that the only decision for a consumer is to buy or not buy. A3' ensures that the consumer value function is only a function of inventory, i.e., $V(I, p) = V(I)$. The lower bound on prices of $\beta^{b-1}\omega_1$ imposed in A3' is necessary for our proof of Proposition 1 to hold when $I = 0$. Assumptions A3' and A4' imply that the stockout cost will be positive, as both $p > 0$ and $\omega_1 \geq 0$. Because of A3', we need to normalize the price coefficient, α, to 1 (we address identification of this parameter in Sections 4 and L). We will relax A3' and A4' in Section L.2 where we explore additional model properties that can help with identification.

Below we formally state three propositions related to a consumer's incentive to purchase. We define next period's inventory as $I'_1(I) \equiv I - 1 + b$ if a purchase is made or $I'_0(I) \equiv \max\{I - 1, 0\}$ if no purchase is made. In order to prove the first two proposition, we first need to derive an individual's optimal policy, which we do in Lemma 2 below.

Lemma 1 *If A1'-A5' hold and $\eta = 0$, then it is optimal to purchase only when $I = 0$, for all β.*

Proof.

Case 1: $\beta = 0$. This is given by A4'.

Case 2: $\beta > 0$.

First note that if a consumer only buys when $I = 0$, then $I \geq 0$. Then by A1' and A4', a consumer always receives $\gamma = 0, \forall t$, she pays the storage cost of a single package, ω_1, for all periods except the period where she starts with a single unit, and she pays the price when she runs out. The consumer will never have more than a single package in storage. Her utility at any level of I (and note that $I \leq b$) will be

$$U(I) = -\frac{1 - \beta^{I-1}}{1 - \beta}\omega_1 \mathbf{1}\{I > 1\} + \beta^I V(0).$$

If the consumer chooses to buy when inventory is zero, it can be shown that the value of having zero inventory, $V(0)$ is

$$V(0) = -\frac{1 - \beta^{b-1}}{(1 - \beta)(1 - \beta^b)}\omega_1 - \frac{p}{1 - \beta^b}. \tag{38}$$

Hence the discounted sum of utility will be

$$U(I) = -\frac{1 - \beta^{I-1}}{1 - \beta}\omega_1 \mathbf{1}\{I > 1\} - \beta^I \frac{1 - \beta^{b-1}}{(1 - \beta)(1 - \beta^b)}\omega_1 - \frac{\beta^I p}{1 - \beta^b}. \tag{39}$$

Claim 2a: It is not optimal for a consumer to choose not to buy when $I = 0$.

If a consumer chooses not to buy when $I = 0$, then the discounted sum of utility received at 0 inventory is $\tilde{V}(0) = \frac{-\nu}{1-\beta}$, and her utility at inventory level I would be

$$\tilde{U}(I) = -\frac{1 - \beta^{I-1}}{1 - \beta}\omega_1 \mathbf{1}\{I > 1\} - \frac{\beta^I \nu}{1 - \beta}.$$

Note that $U(I) > \tilde{U}(I)$, since

$$\frac{1-\beta^{b-1}}{(1-\beta)(1-\beta^b)}\omega_1 - \frac{p}{1-\beta^b} > \frac{1-\beta^{b-1}}{(1-\beta)(1-\beta)}\omega_1 - \frac{p}{1-\beta} > \frac{-\nu}{1-\beta},$$

where the last inequality follows from A4'. This shows Claim 2a.

Claim 2b: It is not optimal for a consumer to choose to buy when $I > 0$.

Consider equation (39). Suppose the consumer considers a one time deviation where she purchases early, and suppose the storage cost does not increase (this will only be the case if $I = 1$, or if $\omega_1 = 0$ and $\omega_2 = 0$, as a result of A5'). Then the consumer pays a price p today and delays the next price paid until $I + b$ periods. However such a strategy cannot be optimal since

$$-p - \frac{\beta^{I+b}}{1-\beta^b}p < -p\frac{\beta^I}{1-\beta^b}.$$

It is clear that if purchasing early increases storage costs, a one time deviation cannot be optimal.

Now last suppose a consumer is using a strategy of always purchasing when $I = \overline{I} > 0$. One can use the same argument to show the consumer could improve her utility by waiting to purchase until $I = 0$: even if storage costs are zero she would improve utility by delaying paying the price p until later on. This proves the claim. ■

Proposition 1 *If A1'-A5' hold and* $\eta = 0$*, then* $\beta * [V(I_1'(I)) - V(I_0'(I))]$ *is strictly decreasing in* I*, for* $\beta > 0, I \geq 0$.

Proof.

Recall that Lemma 2 implies a consumer does not buy until she runs out. As a result, one can derive a formula for $V(I)$, the value function for all inventory values of I, in terms of the value function of having $I = 0$ units, $V(0)$ (the formula for $V(0)$ is equation (38)). We denote as x the number of packages held by a consumer, and $n = I - b(x-1)$ to be the number of units left in the package currently begin consumed by an individual. The formula for the value function will be

$$V(I) = -\frac{1-\beta^{n-1}}{1-\beta}\omega_x - \sum_{k=1}^{x-1}\beta^{n-1+b(k-1)}\frac{1-\beta^b}{1-\beta}\omega_{x-k} + \beta^I V(0) \tag{40}$$

The future value of purchasing is $\beta[V(I-1+b) - V(\max\{I-1,0\})]$. We consider two cases in the proof, and focus on the value function difference $V(I-1+b) - V(\max\{I-1,0\})$. First, suppose that $I > 0$. In this case we will split up the value function difference $V(I-1+b) - V(I-1)$ into two terms:

$$V(I-1+b) - V(I-1) = \Delta_1 + \Delta_2$$

We define Δ_1 to be the difference in the terms in equation (40) that contain storage costs, and let $n = I - 1 - b(x-1)$:

$$\Delta_1 = -\frac{1-\beta^{n-1}}{1-\beta}\Delta\omega_{x+1} - \sum_{k=1}^{x-1}\beta^{n-1+b(k-1)}\frac{1-\beta^b}{1-\beta}\Delta\omega_{x-k+1} - \beta^{n-1+b(x-1)}\frac{1-\beta^b}{1-\beta}\omega_1$$

The second term, Δ_2, is defined to be the difference in the final term of (40), which contains $V(0)$:

$$\begin{aligned}
\Delta_2 &= (\beta^{I+b-1} - \beta^{I-1})V(0) \\
&= \beta^{I-1}(\beta^b - 1)V(0) \\
&= -\frac{\beta^{I-1}(\beta^b-1)}{1-\beta^b}\frac{1-\beta^{b-1}}{1-\beta}\omega_1 - \frac{\beta^{I-1}(\beta^b-1)p}{1-\beta^b} \\
&= \frac{\beta^{I-1}(1-\beta^{b-1})}{1-\beta}\omega_1 + \beta^{I-1}p \qquad (41)
\end{aligned}$$

The derivations above follow from the formula for $V(0)$, which we showed in equation (38). Consider the impact of storage costs on the value function difference. This difference can be written as Δ_1 plus the second term of equation (41):

$$\Delta_1 + \frac{\beta^{I-1}(1-\beta^{b-1})}{1-\beta}\omega_1 = -\frac{1-\beta^{n-1}}{1-\beta}\Delta\omega_{x+1} - \sum_{k=1}^{x-1}\beta^{n-1+b(k-1)}\frac{1-\beta^b}{1-\beta}\Delta\omega_{x-k+1} - \beta^{n-1+b(x-1)}\omega_1$$

As a result the value function difference can be written as

$$V(I-1+b) - V(I-1) = -\frac{1-\beta^{n-1}}{1-\beta}\Delta\omega_{x+1} - \sum_{k=1}^{x-1}\beta^{n-1+b(k-1)}\frac{1-\beta^b}{1-\beta}\Delta\omega_{x-k+1} - \beta^{n-1+b(x-1)}\omega_1 + \beta^{I-1}p. \qquad (42)$$

Note that as I increases, two things happen. First, the storage cost terms in (42) will weakly get more negative due to A5'. Second, the final term, $\beta^{I-1}p$, will decrease since $\beta < 1$. This proves the claim for $I > 0$.

For $I = 0$, it is sufficient to show $V(b) - V(0) - [V(b-1) - V(0)] < 0$, ie, that the value function difference declines when inventory at the beginning of the period decreases from 1 to 0 units. This difference can be written as:

$$\begin{aligned}
V(b) - V(0) - [V(b-1) - V(0)] &= V(b) - V(b-1) \\
&= \omega_1 - \frac{1-\beta^{b-1}}{1-\beta^b}\omega_1 - \frac{1-\beta}{1-\beta^b}p
\end{aligned}$$

The inequality holds as long as $p > \beta^{b-1}\omega_1$, which we assume in A3'.

■

To see the intuition behind the proof of Proposition 1, note that $\beta * [V(I_1'(I)) - V(I_0'(I))]$ captures the net gain in future discounted utility from making a purchase. The higher the current inventory, the smaller this net gain because the point of making a purchase is to avoid paying the stockout cost.

Moreover, because future discounted utility increases with β, the net future gain of making a purchase does as well. This is the main intuition behind the proof of Proposition 2 stated below.

Proposition 2 *If A1'-A5' hold and* $\eta = 0$*, then* $\beta * [V(I_1'(I)) - V(I_0'(I))]$ *is increasing in* β*, for* $I \geq 0$ *and sufficiently small storage costs.*

Proof.

As with Proposition 1, we start with the case where $I > 0$. In this case we need to show

$$\frac{\partial\beta(V(I-1+b) - V(I-1))}{\partial\beta} = V(I-1+b) - V(I-1) + \beta\left(\frac{\partial(V(I-1+b) - V(I-1))}{\partial\beta}\right) > 0. \quad (43)$$

Note that the sign of the first term, $V(I-1+b) - V(I-1) > 0$, depends on the magnitude of storage costs. The derivations in the proof of proposition 1 (see equation (42)) show that this will be positive for sufficiently small storage costs. Next, we want to sign the derivative $\frac{\partial(V(I-1+b)-V(I-1))}{\partial\beta}$. It is possible to show that this derivative is positive if storage costs are sufficiently small.

Note that in general the derivatives of the storage cost component of equation (42) do not have a clear sign. For instance, the derivative of the first term is

$$\frac{(n-1)\beta^{n-2}(1-\beta) - (1-\beta^{n-1})}{(1-\beta)^2}\Delta\omega_{x+1},$$

which could be negative (ie for small β). The terms in the summation sign will have derivatives that look as follows:

$$\left[-(n-1+b(k-1))\beta^{n-2+b(k-1)}\frac{1-\beta^b}{1-\beta} + \beta^{n+b(k-1)}\frac{b\beta^{b-1}(1-\beta) - (1-\beta^b)}{(1-\beta)^2}\right]\Delta\omega_{x-k+1},$$

which again may be negative. The term $-\beta^{n-1+b(x-1)}\omega_1$ will be decreasing in β. Note that if storage costs are zero, then it will be the case that $V(I-1+b) - V(I-1) = \beta^{I-1}p$. In this case the derivative will be

$$V(I-1+b) - V(I-1) + \beta\frac{\partial(V(I-1+b) - V(I-1))}{\partial\beta} = \beta^{I-1}Ip,$$

which is positive. Because the difference in value functions is continuous in storage costs, the above derivative will still be positive, as long as storage costs are sufficiently small.

When $I = 0$, we focus on the expected future value of purchase which is

$$\beta(V(b-1) - V(0)) = \frac{\beta - \beta^b}{1-\beta}\left[1 + \frac{1-\beta^{b-1}}{1-\beta^b}\right]\omega_1 + \frac{\beta-\beta^b}{1-\beta^b}p.$$

The derivative with respect to β of the term $\frac{\beta-\beta^b}{1-\beta^b}$, which multiplies p, is $\frac{1}{1-\beta^b} > 0$. The derivative of the term that multiplies ω_1 has no obvious sign, however if ω_1 is sufficiently small the derivative of $\beta(V(b-1) - V(0))$ will be positive.

■

Note that the size of the storage cost clearly plays a role in the proof of Proposition 2, because if storage costs increase sharply when inventory increases then a forward-looking individual's future payoffs would be reduced by an increase in inventory.

Both Propositions 1 and 2 assume that there is no error term in the utility function. We will derive a third proposition that shows these two propositions still hold in a random utility framework. The argument relies on showing that the expected future value of purchase is continuous in η, and this requires us to make some regularity assumptions on the error term, and put boundedness and sign restrictions on the payoffs:

Assumptions E1-E2

1. Continuity and support: The CDF of the difference in $\varepsilon_1 - \varepsilon_0$, F, is continuous, strictly increasing, and has support $(-\infty, \infty)$.

2. Value function: There exists a bound on η, $\overline{\eta}$, such that if $\eta < \overline{\eta}$, then the following hold:

$$I > 0: \quad -p - (\omega_{B+1} - \omega_B \mathbf{1}\{I > c\} + \beta[V(I+b-1) - V(I-1)] < 0$$
$$I = 0: \quad -p - (\omega_1 - \nu + \beta[V(b-1) - V(0)] > 0$$

   where the number of packages held at the end of the period, B, is $\lceil (I+b-1)/b \rceil$ if $I > 1$ and 0 if $I = 1$.

Lemma 2 *If assumptions A1'-A5' and E1-E2 hold then the expected future value of purchasing, $\beta\left[V(I_1'(I)) - V(I_0'(I))\right]$, is continuous in η.*

Proof.

For $I > 0$, the probability of a purchase can be written as

$$P(I) = F((-p - (\omega_{B+1} - \omega_B \mathbf{1}\{I > 1\} + \beta(V(I + b - 1) - V(I - 1)))/\eta), \tag{44}$$

while the probability of purchase for $I = 0$ is

$$P(I) = F((-p - (\omega_1 - \nu + \beta(V(b - 1) - V(0)))/\eta). \tag{45}$$

The value function for $I > 0$ can be written as

$$V(I) \quad = \quad P(I)(-p - \omega_{B+1} + \beta V(I + b - 1)) + (1 - P(I))(-\omega_B \mathbf{1}\{I > 1\} + \beta V(I - 1)). \tag{46}$$

Under E1 and E2 it is the case that if $I - c \geq 0$ then

$$\lim_{\eta \to 0^-} P(I) = 0,$$

and otherwise

$$\lim_{\eta \to 0^-} P(I) = 1.$$

The limits above are taken from the left as we assume that $\eta \geq 0$. If we consider the first limit, we know that if $I > 0$ then for η sufficiently close to zero, the net value of buying becomes negative, which we assume in E2 (and Lemma 1 implies this inequality holds for $\eta = 0$). For η arbitrarily small and positive the term $(-p - (\omega_{B+1} - \omega_B \mathbf{1}\{I > 1\} + \beta(V(I + b - 1) - V(I - 1)))/\eta$ will be negative and will approach $-\infty$. E1 guarantees that the probability in (44) will approach 0. A similar argument applies to the second limit in the context of equation (45). As a result, it is clear that the limit as η approaches zero of equation (46) will equal the value function that is obtained when $\eta = 0$, which we derive in the proofs of Lemma 1 and Lemma 2. Similar findings will be obtained for the value function when $I = 0$.

■

This Lemma can be used to show the following proposition:

Proposition 3 *If A1'-A5' hold, Propositions 1 and 2 hold for values of η in a neighborhood of 0.*

Proof.

Lemma 1 implies that the value functions are continuous in η, and as a result the value function difference, $\beta\,[V(I_1'(I)) - V(I_0'(I))]$, is also continuous in η. We have shown through Proposition 1 that $\beta\,[V(I_1'(I_1)) - V(I_0'(I_1))] - \beta\,[V(I_1'(I_2)) - V(I_0'(I_2))] < 0$ for $I_1 > I_2$ and $\eta = 0$, and hence continuity implies that this inequality will also hold for values of η in some neighborhood of 0. A similar argument can be applied to the result of Proposition 2: If $\beta_1 [V(I_1'(I); \beta_1) - V(I_0'(I); \beta_1))] - \beta_2 [V(I_1'(I); \beta_2) - V(I_0'(I); \beta_2))] > 0$ for $\beta_1 > \beta_2$ and $\eta = 0$, the inequality will still hold for values of η that are near zero.

■

Proposition 3 enables us to tie Propositions 1 and 2 to observable choice behavior by showing that the propositions hold in a random utility framework. Note that without an error term, myopic and forward-looking individuals behave the same way: they wait until $I = 0$ to make a purchase. However, with an error term an individual may purchase when $I > 0$. Proposition 3 implies that an individual's purchase probability will rise as inventory decreases. Since the magnitude of the incentive to purchase can be measured from choice probabilities, the rate at which the purchase probability drops as inventory rises will be increasing in the discount factor, which we show in Section 4 leads to identification of that parameter.

L Identification with Unobserved Inventory

In this section, we provide additional discussion related to identification when inventory is unobserved and cannot easily be imputed, as we discuss in Section 4. Although inventory itself is unobserved, the time between purchases, which is correlated with inventory, is observed. As a result, the discount factor may be identified from the impact of interpurchase time on a consumer's purchase probability, which is captured by the purchase hazard. We emphasize that the discussion below is based on numerical solutions and is not intended to be a general proof.

In this appendix we relax the assumption of a constant consumption need, since stochastic consumption needs can affect the purchase hazard, and hence potentially interact with identification of other model parameters.[34] In particular, we allow the consumption need c_{it} to be stochastic, and it can only take either 1 or 2. We denote the probability that an individual receives a consumption draw of 1 as π_c. In the numerical solutions below we also assume that the package size, b, is 8 units, we will allow consumers to store up to $M = 3$ packages, and we will assume the error term is standard logit. When we compute the purchase hazard, we will need to simulate out inventory at the time of purchase using the steady state distribution of inventory in the population. To compute

[34] Assuming a constant consumption rate can simplify the identification arguments. In particular, if individuals generally make purchases when they run out of a product, then the consumption rate can be calibrated from the overall quantity an individual buys divided by the length of the time period during which the individual is observed. Then it will be possible to get a very good prediction of inventory from the interpurchase time, and the arguments from the prior section can be applied.

this distribution this we will simulate purchases for 500 individuals for 600 periods. We will assume that in period 0 all individuals have 0 inventory. We find that aggregate inventory appears to reach the steady state at around 50 periods, so we will use periods 400 to 600 to compute steady state inventory. For much of the discussion (except the final discussion of Section L.2) we will also hold prices fixed over time at a level of 2, and will normalize the price coefficient $\alpha = 1$.

Remarks: It is worth comparing our approach here with that in Hendel and Nevo (2006b), which proposes a series of tests for the presence of forward-looking behavior in storable goods markets. Hendel and Nevo (2006b) develop a stockpiling model with endogenous consumption from inventory, and where consumers are able to purchase quantities in continuous amounts. In their setting, a myopic consumer will always purchase exactly the amount she will consume in the period where the purchase occurs, while a forward-looking consumer will purchase for future consumption. An implication of the model developed by Hendel and Nevo (2006b) is that the purchase hazard will be completely flat for myopic individuals, which allows a clean test for the presence of forward-looking behavior. This type of analysis will apply well to settings where consumers have the ability to purchase the product category in small increments: for example, canned tuna or soup.

However, in our setting, we assume package sizes are large relative to consumption needs. Hence, unlike Hendel and Nevo (2006b) the purchase hazard will not be completely flat for myopic individuals in our set up. Because of this complication, we rely on exclusion restrictions to separate out myopic consumers from forward-looking consumers, rather than relying on identification from quantity purchased. Because we rely on exclusion restrictions, we can identify the discount factor in situations where consumers are only able to purchase a single package in a purchase occasion. Another key difference is that in our setting we assume consumption needs are exogenous, in the sense that consumers use just enough of a product to satisfy an exogenous consumption need in each period (for example, one does not get extra utility from consuming more laundry detergent than is needed to do the weekly laundry, or one seldom gets extra utility from drinking more coffee than his/her consumption need). We note that the exclusion restriction may be violated in a setting where consumption is endogenous, since optimal consumption (and hence flow utility) can be a function of inventory.

L.1 Identification with Unobserved Inventory and Zero Storage Costs

We begin by considering the case where $\omega_i = 0$ for $i = 1, ..., M$. In Figure 10, we plot the aggregate probability of purchase in period $t + \tau$ given a purchase in period t for different values of the discount factor. The discount factor primarily affects two features of the purchase hazard. The first feature is the slope of the purchase hazard in the periods immediately after a purchase occurs.

For our particular parameterization, a purchase increases an individual's inventory by 8 units. Since consumption shocks are at most 2 units, it will take someone at least 4 periods to run out and incur a stockout cost. A myopic consumer's flow utility will not change for the first 3 periods after a purchase, and as a result her purchase hazard will be completely flat. This can be seen in Figure 10, where the solid line shows the simulated purchase hazard for a myopic individual. In contrast, for a forward-looking consumer the purchase hazard has a positive slope over the first 3 periods, and this slope increases as the discount factor rises. This occurs because the expected future value of purchase rises as inventory drops, as we showed in Proposition 2. It is notable that without storage costs, there is a clean test for whether individuals are forward-looking or not - if individuals are myopic, the purchase hazard should be flat for the initial few periods after a purchase, provided that it takes some time for individuals to run out of a package after a purchase (which we maintain in Assumption X3). Intuitively, if individuals always use up all their inventory right after a purchase, then it will be difficult to tell if individuals are myopic or not.

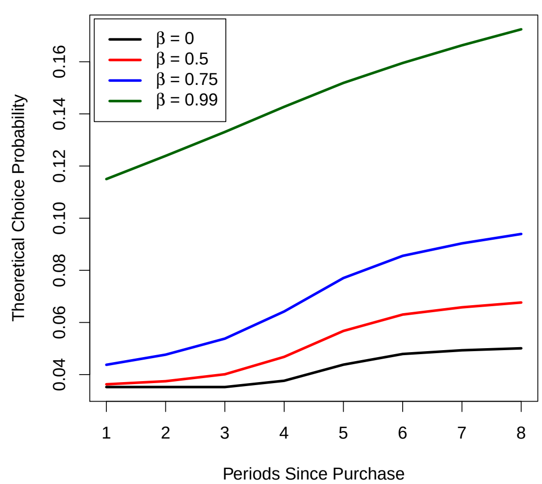

Figure 10: Probability of purchase in period $t+\tau$ given purchase of 1 package in period t. Parameter values $\beta = 0.25$, $\nu = 0.4, \pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \omega_3 = 0, \eta = 1$, $M = 3$, $p = 3.31$, and logit error term.

The second, and more subtle difference between the purchase hazards, is that the purchase hazard becomes smoother as β rises (note that this feature of the purchase hazard is also reflected in the plot of purchase probabilities as a function of observed inventory shown in Figure 6 - for more myopic individuals the choice probabilities jump more as inventory changes). The intuition here is that a myopic consumer is not willing to trade off future utility for current utility, so her

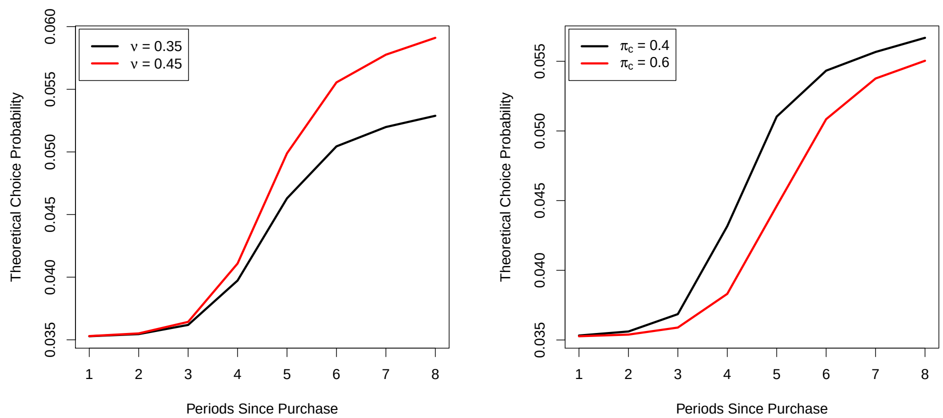

Figure 11: Probability of purchase in period $t+\tau$ given purchase of 1 package in period t, for different values of the stockout cost. Parameter values $\pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \omega_3 = 0, \eta = 1$, $M = 3$, $p = 2$, $\alpha = 1$ and logit error term.

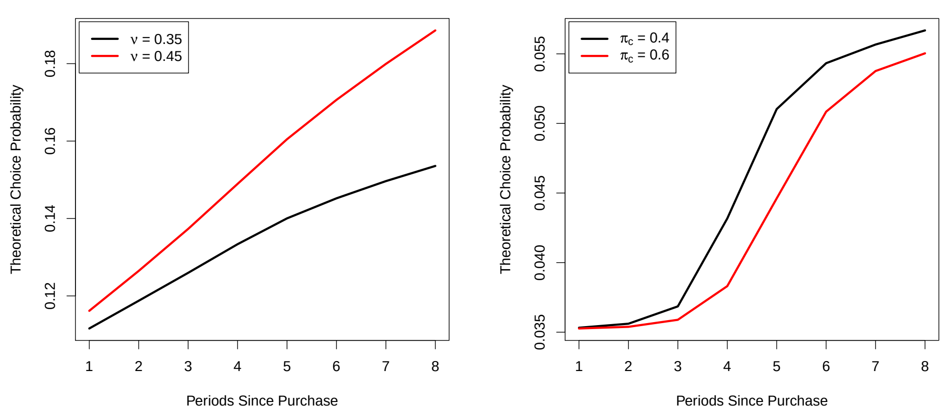

Figure 12: Probability of purchase in period $t+\tau$ given purchase of 1 package in period t, for different values of π_c. Parameter values $\nu = 0.25, \pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \omega_3 = 0, \eta = 1$, $M = 3$, $p = 2$, $\alpha = 1$ and logit error term.

purchase hazard will start to rise sharply at $\tau = 4$, when people in the population start to run out. In contrast, a forward-looking consumer will be more willing to purchase early, and so the purchase hazard will be smoother for such a consumer.

In addition to identifying β, the purchase hazard can be used to identify the other model parameters, ν and π_c. It is relatively easy to see how ν and π_c are separately identified from β, as the two parameters affect the purchase hazard in a very different way. To see this, consider the right panels of Figure 11 and Online Appendix Figure 12, which shows how ν affects the purchase hazard, for low and high values of the discount factor. Most of the impact of a change in ν on the purchase hazard occurs during later rather than earlier periods. This is sensible since ν should have more impact on purchase decisions when consumers begin to run out. Importantly, the shape of the purchase hazard is preserved as ν changes - for the low value of β, the purchase hazard displays a lot of curvature around period 4 for different values of ν.

Similarly, the purchase hazard is very smooth for high values of β, and for different values of ν. The impact of π_c on the purchase hazard for different values of β is shown in the left panels of Figure 11 and Online Appendix Figure 12. For low values of β, the impact of changing π_c is quite similar to that of ν. For high values of β, changing π_c shifts the purchase hazard up and down. We note that the preceding argument suggests that ν and π_c could be difficult to separately identify if β is low. Indeed, as we mentioned in the main body of the paper we encounter this problem in our empirical application in section 5.2, and need to calibrate the consumption need prior to estimating other structural parameters.[35] The price coefficient, α, will shift the overall purchase probability and will simply shift the purchase hazard up or down, and so (in the absence of price variation) it will be identified by the average purchase probability.

L.2 Identification with Unobserved Inventory and Nonzero Storage Costs

If individuals have positive storage costs, the identification argument becomes somewhat more complicated because increases in storage costs can also increase the slope and decrease the curvature of the purchase hazard, similar to the discount factor. To see why, note that when an individual makes a purchase, there is some chance that she has a small amount of a package left over. An individual in this situation will use up the package within a few periods after the purchase, and will observe a decrease in their storage costs. That decrease in storage costs will lead to an increase in the probability of a purchase. To show that this is the case, in Figure 13 we compute the purchase hazards for different discount factors, with the extreme levels of storage costs that are ten times

[35] Even if consumption needs are constant, π_c essentially controls the average consumption need in our simulation, which is $\pi_c + 2(1 - \pi_c)$. A similar identification issue arises even with a constant consumption need.

what we estimate. Although it is difficult to see from the plot, the purchase hazard for a myopic consumer is very slightly positively sloped in the first 3 periods after a purchase - recall in Figure 10 a myopic consumer's purchase hazard was flat in this region, suggesting that the positive storage cost parameter is responsible for the increase in slope. Additionally, the solid line in in Figure 13 is also smoother than the one in Figure 10, suggesting that storage costs can smooth out the purchase hazard.

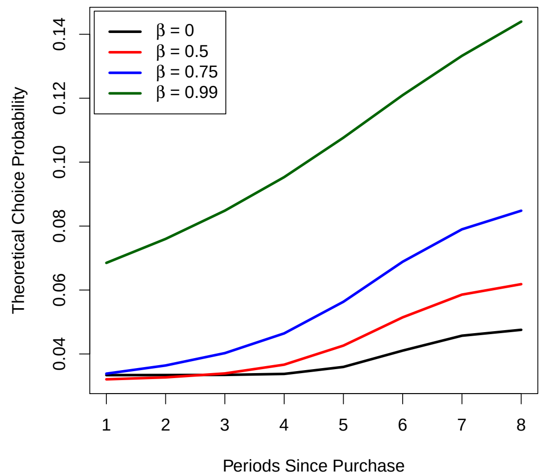

Figure 13: Probability of purchase in period $t+\tau$ given purchase of 1 package in period t. Parameter values $\nu = 0.4, \pi_c = 0.5, \omega_1 = 0.017, \omega_2 = 0.068, \omega_3 = 0.153, \eta = 1,\ M = 3,\ p = 3.31,\ \alpha = 1$ and logit error term.

How do we approach the issue of identification in the presence of storage costs? In the following discussion we explore two different avenues. One avenue is to argue that, because of the exclusion restrictions, there will not be enough storage cost parameters to completely fit the purchase hazard. In our example, we compute the purchase hazard for 8 periods, which means we have at least 8 moments. The number of parameters we have to fit these moments is 7 - three storage cost parameters (ω_1, ω_2,and ω_3), the stockout cost parameter (ν), the discount factor (β), the price coefficient (α), and the probability of a low consumption shock (π_c). Focusing on the discount factor, it is the case that even in the presence of storage costs, an increase in the discount factor still increases the slope of the purchase hazard (at least in early periods) and decreases its curvature. As a result, unless the model's rank condition fails, letting the discount factor be free will provide an improved fit to these features of the purchase hazard.[36] The exclusion restrictions X1 through

[36]i.e., the Jacobian of the model's log-likelihood gradient is singular.

X3 will help to guarantee that this rank condition holds. Exclusion restriction X1 reduces the number of model parameters that one needs to estimate to something manageable. Assumption X3 guarantees that individuals will not run out so quickly that the purchase hazard becomes degenerate. A caveat to the formal approach is that the rank condition may be difficult to verify in practice. Moreover, the identification result will be local, rather than global, as global identification will be more difficult to verify. However, the more overidentifying restrictions a model has, the more likely we can uniquely identify the parameters of the model.

In practice, we can add more restrictions from the moments that are generated by price variation, i.e., the propensity for consumers to stockpile in response to price variation.[37] A forward-looking consumer should become more sensitive to discounts as her inventory drops, since the value of avoiding future stockouts increases with the discount factor. To examine this, we analyze an extension to the model where we allow the price variable to take on two values, 1 and 2, where the value of the price follows a Markov transition process. The probability of the price being 2 given it took a value of 2 in the previous period is 0.8, and the probability of a value of 2 given last period's price was 1 is 0.9. Thus most of the time prices are high, but periodically they drop to the low price for a short time, as is commonly observed in scanner data for storable goods. Additionally, we relax the restriction that individuals can only purchase a single package, and allow individuals to purchase up to 2 packages at once. Because we assume that the transition process for prices is known to consumers, individuals may stock up when prices are low.

One measure of an individual's propensity to stock up is the amount by which the probability a person buys 2 units relative to 1 unit increases when the price drops. Intuitively, if a consumer is myopic she has no need to purchase more than a single package - as a result, the propensity to purchase more than one package should be driven entirely by the error distribution and any potential change in storage costs. However, a forward-looking consumer should become relatively more likely to purchase multiple units at low prices, and this likelihood should increase as inventory drops (for sufficiently low values of inventory).

In Figure 14, we plot the ratio of probability of buying 2 units to 1 unit at the low price minus the same ratio at the high price, given t periods have elapsed since the last purchase occured. The left panel shows how this probability difference changes if there are no storage costs. It is notable that for a myopic individual, the propensity to stockpile in response to deals is completely unaffected by inventory, since the moment we show in the graph is totally flat (This occurs because if the error term is logit, $\beta = 0$, and storage costs are zero, for a myopic individual this moment will be $\exp(2\alpha) - \exp(\alpha)$; The stockout cost drops out of the probability ratio used to compute this

[37]Model moments due to price variation are commonly used in estimating stockpiling models.

moment). However, this moment increases as inventory drops for forward-looking individuals, and the slope of the curve rises with larger values of β.

The right panel shows how the same moment changes with interpurchase time, with positive storage costs. Here, the propensity to stockpile in response to low prices still rises if inventory is sufficiently low. An important point to note is that if the discount factor is low, the slope of this moment is relatively unaffected by the storage cost (this can be seen by comparing the curves corresponding to $\beta = 0$ and $\beta = 0.5$ on the left panel to those on the right panel). However, if the discount factor is high, the storage costs decrease the slope of the moment, which is intuitive - if storage costs are high and individuals are forward-looking they should have less incentive to stockpile at low prices. The fact that the storage costs and discount factor have the opposite effect on the propensity to stockpile, while they both increase the slope of the purchase hazard, will help us to separately identify them. We note that price variation will also identify the price coefficient, but the price coefficient can be identified from the average change in the purchase probability for the low versus the high price; the preceding argument relies on how stockpiling in response to deal sensitivity changes as inventory changes.

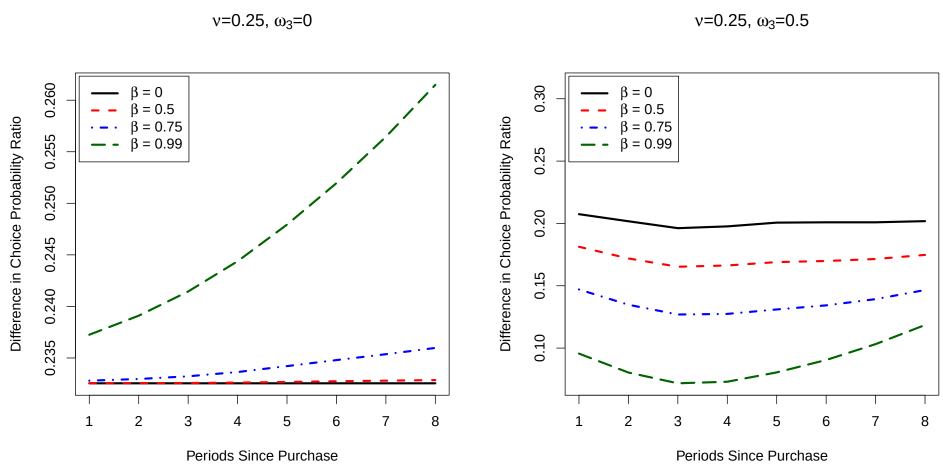

Figure 14: Ratio of probability of buying 2 units to 1 unit at $p = 1$ minus the same ratio at $p = 2$ in period t, given purchase of 1 package in period 0. Parameter values $\nu = 0.25, \pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \omega_3 = 0$ or $0.5, \eta = 1$, $M = 3$, and logit error term.

L.3 Artificial Data Experiments

To provide further evidence that the model above can be identified in realistic settings, even if inventory is unobserved to the researcher, we perform a series of artificial data experiments where we simulate data and recover the underlying parameters. Our simulated data is made up of of

500 households whose purchases are simulated from our model for 600 periods. When simulating purchases, in period 1 all individuals start with 0 inventory. In field data consumers will have been making purchases prior to the beginning of the data collection; to mimic such a situation, we only use periods 201 to 600 to recover the model parameters. We estimate the model parameters using simulated maximum likelihood, where the likelihood is taken over quantities purchased - individual inventories are assumed to be unobserved during estimation, as is the case in field data. As a result, inventories must be simulated out when constructing the likelihood. To integrate out unobserved inventory in the likelihood construction, we set individual inventories to 0 in period 201, and draw a series of consumption shocks for each consumer for periods 201 through 600. We draw 100 such consumption paths for each consumer. With simulated consumption shocks and observed purchase quantities one can construct an estimate of inventory in each period. To mitigate the impact of assuming 0 inventory in period 201 on the estimated parameters, the purchase likelihood is constructed using purchases from periods 401 through 600, and for each individual we average over the 100 simulated paths to reduce simulation noise. Our procedure of using the first part of a sample to construct inventories is standard in the literature (Erdem, Imai, and Keane (2003), Hendel and Nevo (2006a)).

To estimate the price coefficient we need sufficient price variation. We assume that the price takes on 3 values, 0.5, 1 and 2, and prices follow a Markov transition process. If the price is 0.5 or 1, it stays the same with probability 0.1, and becomes 2 with probability 0.9. If the price is 2, it stays 2 with probability 0.8 and transits to 0.5 or 1 with probability 0.1. The model specification is derived from the theoretical model of Section 3: We allow consumers to purchase 2 packages at most, set the package size $b = 8$, set consumption shocks to be in the set $\{1, 2\}$, and restrict consumers to hold at most 3 packages. The error term is assumed to be logit and the weight on it is set to $\eta = 1$.

The results of the artificial data experiment are shown in Table 26. The top panel shows how the parameter identification is affected by including storage costs and by letting the storage cost function be more flexible. In the first 3 columns of this panel, we estimate the model in a situation where storage costs are zero. The first column shows the estimated parameters, the second the standard errors, and the third is the true values of the parameters. All the parameters are well identified. The next three columns show how the results change if we allow ω_3 to be free, while holding ω_1 and ω_2 fixed at 0. The parameter estimates are still close to the truth, although the standard errors are larger. If we allow all 3 storage cost parameters to be positive, and estimate all of them, the standard errors rise significantly, although the parameter estimates are still relatively close to the truth.[38] Note that the precision on the discount factor drops as the number of storage cost parameters increases. This highlights the importance of exclusion restrictions. They allow for more precise identification of the discount factor when it is applied to more of the state space: For instance, in cases where it might be reasonable to assume that the cost of storing the first 1 or 2 packages is 0. Turning to the other parameters, the storage cost coefficients are also imprecisely estimated, especially ω_3 - its standard error is three times higher than the estimate from the situation where ω_1 and ω_2 are fixed. This also highlights the fact that assuming zero storage costs for the first few packages may aid identification. It is notable that all the other model parameters, such as α, ν and π_c, are well-identified even if storage costs are flexible.

The bottom panel of the table shows how identification of the discount factor varies as consumers get more forward-looking. The first column shows the case where consumers are essentially myopic. In this case, the discount factor is not precisely identified. The reason for this is that a consumer with a positive, but low discount factor such behaves very similarly to a myopic consumer. As the discount factor rises, the precision with which we can estimate it also rises.

[38] We do not show the results when we allow for 2 storage costs to be free to save space; in that case the standard errors are a little higher than when we have only a single free storage cost parameter.

Table 26: Artificial Data Experiment: Results

Parameter	No Storage Costs Est	S.E.	Truth	ω_2 Free Est	S.E.	Truth	ω_1, ω_2 Free Est	S.E.	Truth
Price Coeff (α)	1.004	0.007	1	1.002	0.014	1	1.001	0.015	1
Stockout Cost (ν)	0.098	0.01	0.1	0.103	0.006	0.1	0.101	0.036	0.1
Discount Factor (β)	0.957	0.016	0.95	0.957	0.032	0.95	0.954	0.067	0.95
ω_1	-	-	0	-	-	0	0.105	0.059	0.1
ω_2	-	-	0	-	-	0	0.243	0.051	0.25
ω_3	-	-	0	0.499	0.055	0.5	0.508	0.142	0.5
π_c	0.489	0.007	0.5	0.49	4.72e-04	0.5	0.5	0.002	0.5

Parameter	$\beta = 0.001$ Est	S.E.	Truth	$\beta = 0.6$ Est	S.E.	Truth	$\beta = 0.99$ Est	S.E.	Truth
Price Coeff (α)	1.002	0.007	1	1.002	0.009	1	1	0.014	1
Stockout Cost (ν)	0.096	0.022	0.1	0.1	0.011	0.1	0.102	0.006	0.1
Discount Factor (β)	0.001	0.149	0.001	0.619	0.052	0.6	0.994	0.034	0.99
ω_1	-	-	0	-	-	0	-	-	0
ω_2	-	-	0	-	-	0	-	-	0
ω_3	0.479	0.061	0.5	0.47	0.044	0.5	0.488	0.059	0.5
π_c	0.492	0.001	0.5	0.496	2.80e-04	0.5	0.494	4.03e-04	0.5

M Figures

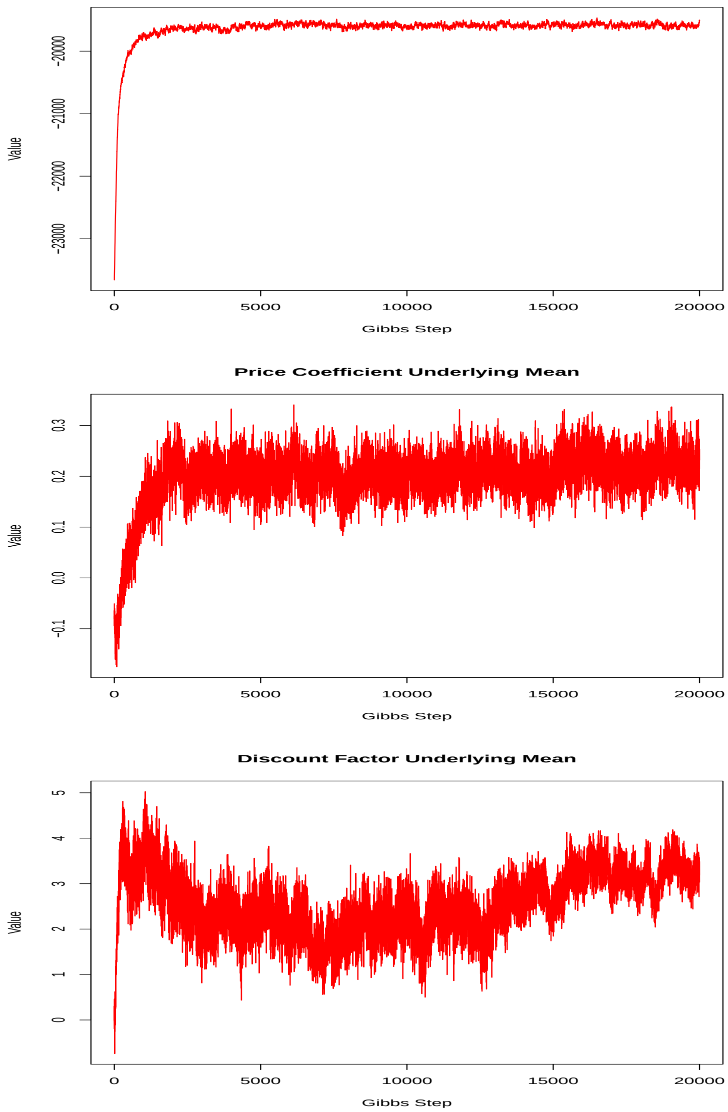

Figure 15: Appendix Figure: Plots of Log-Likelihood (top), MCMC Draws for Selected Dynamic Parameters