

Global Research Unit Working Paper #2019-023

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

Andrew T. Ching, Johns Hopkins University
Matthew Osborne, University of Toronto

© 2019 by Ching & Osborne. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

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

Andrew T. Ching[†] and Matthew Osborne[‡]

Abstract

Understanding how forward-looking consumers respond to price promotions in storable goods markets is an important area of research in empirical marketing and industrial organization. In prior work, researchers have assumed that consumers in these markets are very forward-looking, and calibrated their weekly discount factors to levels around 0.9995. This calibration has been used because earlier research has assumed that a consumer's storage cost is a continuous function of inventory, which rules out exclusion restrictions that can be used to identify the discount factor. We show that by properly modeling 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), *natural exclusion restrictions* arise that allow for the discount factor to be point identified. In an application to a storable good category, we find that weekly discount factors are very heterogeneous across consumers, and are on average 0.71. We show through a counterfactual exercise that if one used a model which fixed the discount factor to be consistent with the standard calibrated value, one would overpredict the effect of increased promotional depth for a product on its quantity sold by 18% in the short-term, and 15% in the long-term.

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

*The authors thank the editor, the associate editor and the two anonymous referees for their helpful comments. The authors also thank Øystein Daljord, Aviv Nevo, Stephan Seiler, Steve Stern, Chenyu Yang, Xinyu Zhang and seminar participants at the Summer Institute for Competitive Strategy, the Stanford Institute for Theoretical Economics, Washington DC Industrial Organization Conference, Boston College, McMaster University, University of Toronto, University of Western Ontario, The Wharton School, Purdue University, McGill University, University of Hong Kong, Fudan University, Stony Brook University and University of Rochester for providing helpful feedback on earlier drafts.

[†]Carey Business School, Johns Hopkins University, Baltimore, Maryland 21202. email: andrew.ching@jhu.edu.

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

1 Introduction

Consumer stockpiling behavior of packaged goods has been widely studied in both the empirical marketing (Erdem, Imai, and Keane (2003), Sun (2005), Chan, Narasimhan, and Zhang (2008), Seiler (2013), Liu and Balachander (2014), Haviv (2014), Osborne (2018b)) and empirical industrial organization literatures (Pesendorfer (2002), Hendel and Nevo (2006), Hendel and Nevo (2013), Wang (2015), Pires (2016), Osborne (2018a)). Prices often follow a “hi-lo” pattern where they take on a low promotional value for one or two weeks, and return to a higher regular retail price for a longer period of time. When facing such a price process, forward-looking consumers will behave strategically and respond to temporary price promotions by stockpiling the good, as they understand that prices will be high in the future and they wish to have the good available for later use. Much of the empirical literature on this topic employs the discrete choice dynamic programming framework (Rust 1994), where a dynamic structural model of consumer behavior is estimated on consumer purchase data and used to make counterfactual predictions. In the marketing literature, past research has used structural models to quantify the impact of changing the price process’ variance, promotional frequency, and promotional depth on brand and category sales (Erdem, Imai, and Keane (2003), Liu and Balachander (2014) and Osborne (2018b)). Understanding demand responses to changes in the price process is of critical importance to retail managers who wish to optimally allocate limited promotional budgets. In the empirical industrial organization literature, structural models have been used to understand the impact of stockpiling on merger analysis (Hendel and Nevo 2006), food taxation (Wang 2015), and price index construction (Osborne 2018a).

A key parameter in dynamic discrete choice models is the *discount factor*, which measures the strength of forward-looking behavior. The closer the discount factor is to 1, the more weight consumers put on future payoffs when making current decisions. However, none of the preceding dynamic structural papers on stockpiling estimate the discount factor. Instead, they assume that consumers make inter-temporal trade-offs based on the prevailing interest rate, calibrating the discount factor accordingly. This calibration approach leads to a weekly discount factor of about 0.9995.¹ The calibrated discount factor 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

¹The calibration approach assumes a yearly interest rate of 5%, which is consistent with U.S. real interest rates in the period 2001-2010. This interest rate implies a yearly discount factor of 0.95, and would produce 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 calibration implies to reduce the computational burden of estimating the dynamic model. For example, Seiler (2013) uses a value of 0.998.

be context specific. Moreover, in stated choice experiments performed by Dubé, Hitsch, and Jindal (2014), consumers appear to be much less forward-looking than what economic theory implies, with average annual discount rates of 0.43. Dubé, Hitsch, and Jindal (2014) and Yang and Ching (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 restrictions on the structure of the current period utility function, the discount factor cannot be identified (see Rust (1994), Magnac and Thesmar (2002)). Recently, Magnac and Thesmar (2002), Fang and Wang (2015), and Abbring and Daljord (2018) have examined the question of whether the discount factor can be identified when a dynamic model has exclusion restrictions. 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 this 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, most recent research still fixes the discount factor according to the interest rate, because the state variables of these dynamic models do not provide exclusion restrictions. Moreover, even with such restrictions, point identification is not guaranteed (Abbring and Daljord 2018).

Our paper makes two important contributions to the literature on stockpiling. First, we argue that one of the key state variables in these models, inventory, provides *natural exclusion restrictions*; we show that with these restrictions, all model parameters, including the discount factor, are point identified.² Our key insight is that if an individual’s storage cost is a function of the number of packages held, then 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 in her home, and her consumption rate is constant over time.³ 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

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

³For example, this consumer has a habit of wearing clean clothes every day, and as a result does a single load of laundry every week.

of inventory remaining, the more forward-looking a consumer is, the more intense this feeling of urgency will get. In contrast, the existing literature has assumed that a consumer’s storage cost is a continuous function of inventory, which rules out these exclusion restrictions that naturally arise from the institutional features of the problem. With a continuous and increasing storage cost function, a consumer’s purchase probability will increase smoothly as inventory drops, even if the consumer is myopic, making it difficult to identify forward-looking behavior.

To implement this insight, we make an additional assumption that an individual’s consumption rate is constant (but it can differ across individuals).⁴ This assumption, together with the standard approach to handle the initial value of one’s inventory level, allows us to compute individual specific inventory at any point in time. These assumptions allow us to take advantage of exclusion restrictions and reduce the computational burden of estimating the model.

Our second contribution relates to the empirical and managerial implications of estimating the discount factor. We estimate a dynamic structural model of consumer stockpiling behavior on scanner data for laundry detergents, and recover the population distribution of discount factors. We find that the average weekly discount factor is 0.71, which is significantly lower than the calibrated benchmark. Additionally, there is a significant amount of heterogeneity in discount factors, as the 33rd and 66th percentiles of the population distribution are 0.62 and 0.99, respectively.

We demonstrate that estimating the discount factor matters through a series of counterfactual exercises, which compare the impact of estimating versus calibrating the discount factor on predictions of how consumer demand for a product responds to changes in the product’s price process. Often times, firms need to evaluate a new price promotion policy soon after its implementation. Consequently, consumer expectations about the price process may not have adjusted in a short time frame. Therefore, we construct our comparisons for two scenarios: (i) the short-term, where consumer price expectations correspond to the actual process observed in the data; and (ii) the long-term, where price expectations correspond to the counterfactual price process. We evaluate two different counterfactual changes to the price process: one where we increase the depth of price promotions for a popular product, and the other where we increase their frequency.

In the model where the discount factor is estimated, we predict that by offering deeper discounts, one can significantly increase consumer demand in both the short- and long-term. Relative to these predictions, the model with the calibrated discount factor of 0.9995 significantly overpredicts quantity sold, by 18% in the short-term, and 15% in the long-term. When the discount factor is calibrated at 0.9995, consumers are more responsive to increased promotional depth because more

⁴This assumption is motivated by our observation that the consumption rate should be fairly constant for some product categories (e.g., laundry detergent).

forward-looking consumers have a stronger incentive to stockpile the product on promotion, instead of other products. This effect is amplified in the short-term because consumers underestimate the likelihood that such a promotion will occur in the future, leading to a larger bias in the short-term prediction. We find that the calibrated discount factor model also overpredicts the impact of increased promotional frequency on quantity sold, but by a smaller amount. These counterfactual results suggest that if exclusion restrictions are available to help identify the discount factor, it should be estimated, rather than calibrated.

The rest of the paper is presented as follows. In Section 2, we discuss related work. In Section 3 we use a stylized model to present the exclusion restrictions, and Section 4 develops our identification proof. We describe our empirical application in Section 5, the counterfactuals in Section 6 and discuss future research paths in Section 7.

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 are observed to the researcher, and there is no unobserved heterogeneity across consumers. In this setting, under a set of regularity conditions on the error term, one can non-parametrically 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, the moments that identify the model parameters are the probabilities 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 type of 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 it. 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.⁵ In our setting, we show that we can obtain point identification if we exploit multiple exclusion restrictions.

To our knowledge, there are only a handful papers that explore such an identification argument to estimate a consumer’s discount factor or her incentive to consider future payoffs (Chevalier and Goolsbee (2009), Ishihara and Ching (2012), Chung, Steenburgh, and Sudhir (2013), Lee (2013), Ching, Erdem, and Keane (2014), Ching and Ishihara (2018), Dalton, Gowrisankaran, and Town (2019), De Groot and Verboven (2019)). These papers investigate sales force compensation schemes, rewards programs, consumer learning, 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), Sun (2005), Hendel and Nevo (2006), Chan, Narasimhan, and Zhang (2008), Seiler (2013), Liu and Balachander (2014), Pires (2016)). 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.⁶ 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.

In on-going research, 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

⁵Fang 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.

⁶Note 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.

use brands within inventory in a last-in-last-out order and consumption rates are constant, the discount factor can be identified. Their approach is much more computationally demanding than ours because it requires consumers to track inventory at the brand level. Since most field datasets have many brands, the state space of such a model would become very large. Moreover, our approach does not rely on any particular assumption related to the order of consumption of brands within inventory.

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) then 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.⁷

3 Exclusion Restrictions in the Stockpiling Model

In this section we develop a simple dynamic model of consumer purchase behavior for a storable good, which we use to demonstrate how our exclusion restrictions can be used to identify the model’s parameters. This model abstracts away from many complications we include in our empirical model, such as price variation, preference heterogeneity, and purchase of multiple packages at once. We assume that the researcher observes a market containing N consumers making purchase decisions over T periods. Consumers are forward-looking and discount the future at a rate $\beta < 1$. In this stylized model, we assume that a single product is available to consumers in a single 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), an exogenous consumption need (c), and a choice-specific error (ε_{ijt}). In each period, the consumer’s decision is whether or not to purchase a single package of the product: $j \in \{0, 1\}$. After making her purchase, the consumer enters her consumption phase and consumes an amount c^* , which is determined by her consumption need and available inventory. We assume that the

⁷Ching, Erdem, and Keane (2014) extends this line of research to structural learning models by taking advantage of the exclusion restrictions implied by the Bayesian updating rule. As a result, even though Ching, Erdem, and Keane (2014) did not observe an individual’s current payoffs, they are able to recover an individual’s expected future payoffs.

maximum amount of inventory an individual can hold is $M \cdot b$, where b is the size (or volume) of a package, and M is the maximum number of packages that can be stored. For simplicity of exposition, we will assume that b and c are both strictly positive integers, and b is divisible by c . At the end of the period, the consumer incurs a storage cost $s(\cdot; \omega)$. ω is a vector of parameters determining how storage costs vary with the number of packages held. We make three modeling assumptions which will guarantee that inventory generates exclusion restrictions, and allow identification of model parameters from consumer choice probabilities.

Modeling Assumptions Related to Exclusion Restrictions

- X1 Consumption $c^* = c$ if $I + bj \geq c$, and $c^* = 0$ otherwise. Flow utility from $c^* \geq c$ is equal to a constant, $\gamma \geq 0$. If $c^* < c$, the individual incurs a stockout cost of $\nu > 0$.
- X2 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 .
- X3 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: $c \leq b/3, M \geq 2$.

Assumption X1 says that consumption rates will be constant over time and equal to c (if end of period inventory is at least c). This implies that the individual's end of period inventory will be $I_{i,t+1} = \min\{\max\{I_{it} + b \cdot j - c, 0\}, M \cdot b\}$. X1 additionally states that underlying consumption utility is equal to $-\nu < 0$ if consumption is smaller than c , and is equal to γ for consumption levels greater than or equal to c . We assume that $\gamma \geq 0$ and $\nu > 0$ so an individual would prefer to avoid stockouts, and hence stockpile even with constant prices. With varying prices, individuals would have an incentive to stockpile during price promotions, even without a stockout cost. This assumption is slightly stronger than necessary: Because either of these parameters can be normalized, it is only necessary that $\gamma > -\nu$. We note that assumption X1, along with the assumption that b is divisible by c , will imply that a consumer will never have positive inventory below c . The divisibility assumption is made for simplicity of presentation here.

Assumption X2 makes it possible for there to be some values of I for which $s(\cdot; \omega)$ does not change. Together, Assumptions X1 and X2 allow for the possibility that there are values of I for which exclusion restrictions can hold. Note that the number of packages held at the end of a period can be written as the following function of inventory: $B_{i,t+1}(j, I, c) = \lceil \max\{(I_{it} + b \cdot j - c)/b, 0\} \rceil$.⁸ For the remainder of this section and Section 4, we allow the storage cost function to be non-

⁸The ceiling function $\lceil \cdot \rceil$ returns the smallest integer that is greater than or equal to its argument

parametric:

$$s(B; \boldsymbol{\omega}) = \omega_B. \quad (1)$$

We assume that the cost of storing 0 packages is 0.

Assumption X3 guarantees that there will exist some levels of inventory where the exclusion restrictions hold, by ensuring that for most inventory levels consumers face, the number of packages held will remain unchanged (and hence storage costs will also stay the same). In particular, we show in Section 4 that in order to uniquely identify β , it is necessary to observe choice probabilities for at least three levels of inventory where the exclusion restrictions hold.

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

$$u(j, I_{it}, \varepsilon_{ijt}, p, c; \boldsymbol{\theta}) = \begin{cases} \gamma - s(B_{i,t+1}(j, I, c); \boldsymbol{\omega}) - \alpha \cdot p \cdot j + \varepsilon_{ijt} & \text{if } I_{it} + b \cdot j \geq c \\ -\nu - \alpha \cdot p \cdot j + \varepsilon_{ijt} & \text{otherwise} \end{cases}, \quad (2)$$

where $\boldsymbol{\theta} = (\alpha, \beta, \gamma, \nu, \boldsymbol{\omega})$ is a vector of the consumer utility coefficients and the discount factor, α is the price coefficient, and ε_{ijt} is a choice-specific error. γ and ν cannot be separately identified, and hence, we normalize $\gamma = 0$.⁹ Denoting the vector of choice-specific errors as $\boldsymbol{\varepsilon}_{it}$, the consumer's Bellman equation can be written as follows:

$$V(I_{it}) = E_{\boldsymbol{\varepsilon}_{it}} \max_{j \in \{0,1\}} \{u(j, I_{it}, \varepsilon_{ijt}, p, c; \boldsymbol{\theta}) + \beta V(I_{i,t+1})\}. \quad (3)$$

Before turning to identification, we make some comments on our modeling assumptions. First, in practice one may consider imposing a functional form on storage costs s . One possibility is that the storage cost could be close to zero up to some limit, and increase dramatically beyond that. For instance, for many households the laundry room may have reserved space for a few bottles of detergent. Assumptions X1-X3 will limit the types of product categories where our identification approach can be applied. In particular, they apply to product categories such as laundry detergent, cleaning products, or breakfast cereals. In these categories, it is arguable that individual consumption rates are relatively constant: consistent with X1 and X3, people likely eat a fixed amount of cereal for breakfast and do laundry at regular intervals, and the consumption rates in these categories are low relative to the size of a package. Consistent with X2, the storage cost of the product will correspond to the space taken up by the bottle or package, but not the amount of the product within the package.

⁹It can be shown that if we reduce γ by $\epsilon > 0$ and add ϵ to ν , the difference in choice-specific utility remains unchanged at all inventory states.

Assumption X1 is likely to be violated for product categories such as snacks or sweets where the temptation to consume may rise with larger inventory stocks (e.g., Sun (2005), Chan, Narasimhan, and Zhang (2008)), and X2 is likely violated for products such as potato chips where the package size can be shrunk as the product is used up. X3 would likely be violated for product categories where every time consumption occurs, a package is used up, such as canned tuna or canned soup.

4 Theoretical Identification Proof

In this section, we present a constructive proof that the parameters of the model presented in Section 3 can be uniquely identified from choice data. We make three additional modeling assumptions below:

Modeling Assumptions

A1 The consumption rate $c = 1$.

A2 The choice-specific error term, ε_{ijt} , follows a type-1 extreme value distribution.

A3 Inventory, I_{it} , is observed to the researcher.

Assumption A1 is a normalization we make for notational simplicity. Assumption A2 is a standard assumption that is maintained by the previous literature (Fang and Wang (2015), Abbring and Daljord (2018)), which makes the derivations of choice probabilities simpler. Identification can still be obtained with a non-logit error that has an increasing CDF and full support on the real line. We maintain A3 because our constant consumption rate assumption, together with an assumption on initial inventory, allows us to compute an individual's inventory at any point in time. In this section of the paper, we assume the consumption rate is known, and in our empirical application we estimate the consumption rate outside the model. We provide discussion related to how we estimate the consumption rate in the empirical application in Section 5.1. We provide an informal discussion of how identification may be possible if inventory is unobserved to the researcher in Appendix A.

Below, we define a rank condition, R1, that is necessary for identification of β . In Section 5.3, we provide evidence that the purchase behavior observed in our data set satisfies R1.

Rank Condition R1

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

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

Theorem 1 *Suppose Assumptions A1-A3 (modeling assumptions), X1-X3 (exclusion restrictions) and R1 (rank condition) hold. Then the parameters of the stockpiling model, α , β , ν , and ω_1 through ω_M , are uniquely identified.*

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

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

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. We leave out some details on derivations, which are available from the authors upon request.

Without loss of generality, let us assume that the exclusion restrictions hold for inventory levels I through $I + 2$, where $I + 2 \leq b + 1$ and $I > 1$. Additionally, suppose that Rank condition R1 holds at I . It is straightforward to show that the condition that $\hat{P}(I) \neq \hat{P}(I + 1)$ in rank condition R1 implies that $\beta > 0$.¹⁰ If we define

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

and $\Delta v(I) = v_1(I; \boldsymbol{\theta}) - v_0(I; \boldsymbol{\theta})$, then for $\beta > 0$ it can be shown that if the exclusion restrictions hold for inventory levels I and $I + 1$, then

¹⁰The case of $\beta = 0$ is straightforward. If choice probabilities are flat for all inventory levels belonging to the same package, we can infer that $\beta = 0$.

$$\Delta v(I) = \frac{1}{\beta} \Delta v(I+1) - \hat{\Phi}(I) - \frac{2\beta-1}{\beta}(\omega_2 - \omega_1) - \frac{\beta-1}{\beta} \alpha p.$$

If the above equation holds for three levels of inventory, I through $I+2$, it is possible to difference out the terms $(\omega_2 - \omega_1)$ and αp to express β as

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

for a value of $I > 1$ 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 if the expression in 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:

$$\hat{\alpha} = -\frac{\Delta \log(\hat{P}(Mb))}{p}.^{11}$$

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

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

where the terms h_0 and h_1 are functions of β and choice probabilities, \hat{P} , for inventory values between 0 and b :

$$\begin{aligned} h_0(\hat{\beta}) &= \frac{\hat{\beta}^{b-1}}{1-\hat{\beta}} \ln \left(1 + \frac{\hat{P}(0)}{1-\hat{P}(0)} \right) + \sum_{i=0}^{b-2} \hat{\beta}^i \ln \left(1 + \frac{\hat{P}(b-1-i)}{1-\hat{P}(b-1-i)} \right) - \frac{1}{1-\hat{\beta}} \ln \left(1 + \frac{\hat{P}(0)}{1-\hat{P}(0)} \right), \\ h_1(\hat{\beta}) &= \frac{\hat{\beta}^b}{1-\hat{\beta}} \ln \left(1 + \frac{\hat{P}(0)}{1-\hat{P}(0)} \right) + \sum_{i=0}^{b-1} \hat{\beta}^i \ln \left(1 + \frac{\hat{P}(b-i)}{1-\hat{P}(b-i)} \right) - \frac{1}{1-\hat{\beta}} \ln \left(1 + \frac{\hat{P}(0)}{1-\hat{P}(0)} \right). \end{aligned}$$

Note that the system of equations (8) 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

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

¹¹We 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 being used and sets her inventory level to $M \cdot b$. 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.

where $V(b+1)$ and $V(1)$ can be expressed in terms of choice probabilities and the parameters $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\omega}_1$:

$$\begin{aligned}
V(1) &= \frac{-\hat{\beta}\hat{\nu}}{1-\hat{\beta}} + \frac{\hat{\beta}}{1-\hat{\beta}} \ln \left(1 + \frac{\hat{P}(0)}{1-\hat{P}(0)} \right) + \ln \left(1 + \frac{\hat{P}(1)}{1-\hat{P}(1)} \right), \\
V(I) &= -\frac{1-\hat{\beta}^{I-1}}{1-\hat{\beta}}\hat{\omega}_1 - \frac{\hat{\beta}^I\hat{\nu}}{1-\hat{\beta}} + \frac{\hat{\beta}^I}{1-\hat{\beta}} \ln \left(1 + \frac{\hat{P}(0)}{1-\hat{P}(0)} \right) + \sum_{i=0}^{I-1} \hat{\beta}^i \ln \left(1 + \frac{\hat{P}(I-i)}{1-\hat{P}(I-i)} \right), \\
&\text{for } 2 \leq I \leq b+1.
\end{aligned}$$

Higher values of the storage cost can be derived from the equation

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

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: $\hat{\alpha}$, $\hat{\beta}$, $\hat{\nu}$, $\hat{\omega}_1$, ..., $\hat{\omega}_{B-1}$.

■

The proof above builds on prior work by Ching and Ishihara (2018), who consider a simple dynamic store choice problem with rewards programs, but is much more involved than theirs. They are able to take advantage of the feature that the value of state variables (rewards points at stores) 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. Rossi (2018) applies the identification results in Ching and Ishihara (2018) to estimate the consumer discount factor using data from a retail gasoline reward program.

4.1 Discussion of the Intuition Behind Identification

In this section, we provide some additional comments on equation (7) in Theorem 1, which defines β . A theoretical contribution of our research is to show that it is possible to uniquely identify β from choice data in the stockpiling model using exclusion restrictions. This is important because Abbring and Daljord (2018) have recently shown that even if one imposes the minimal exclusion restrictions proposed by Fang and Wang (2015), β is generally not point-identified. In our case, in order to express β only in terms of choice probabilities, we must solve two moment conditions,

$$\Delta \log(\hat{P}(I+1)) - \Delta \log(\hat{P}(I)) = \beta(V(I+b) - V(I) - (V(I+b-1) - V(I-1))) \quad (11)$$

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

where the exclusion restrictions hold for $I, I + 1$, and $I + 2$. We emphasize that the value function differences such as $V(I+b) - V(I) - (V(I+b-1) - V(I-1))$ are functions of β and choice probabilities, in addition to inventory. We have suppressed the additional arguments of V for simplicity of presentation. In the standard formulation of the exclusion restrictions, one would invert one of the above equations to obtain β . To illustrate how an identification failure can occur with only a single moment restriction, in Figure 1 we plot the right hand side of equation (11) using a solid line and equation (12) using a dotted line, for $I = 3$. The left hand sides of these two equations are data that can be estimated by the researcher. Suppose that the researcher were to only use equation (11) for identification, and the left hand side was $\Delta \log(\hat{P}(4)) - \Delta \log(\hat{P}(3)) = -0.14$. There are two possible values of β that are solutions to equation (11), at 0.79 and 0.95. However, if the researcher were to also exploit equation (12), and the left side of this equation were $\Delta \log(\hat{P}(5)) - \Delta \log(\hat{P}(4)) = -0.10$, then only $\beta = 0.79$ would be a solution to both moment equations.

We note that the proof of identification describes the minimum number of moments one would need to identify the discount factor, as well as the other model parameters. In practice, the number of moments available to the researcher would likely be much larger. In many product categories, it may take consumers many periods to use up a package, meaning that the exclusion restrictions should hold for many values of inventory. These additional moments will improve the precision of the model parameter estimates.

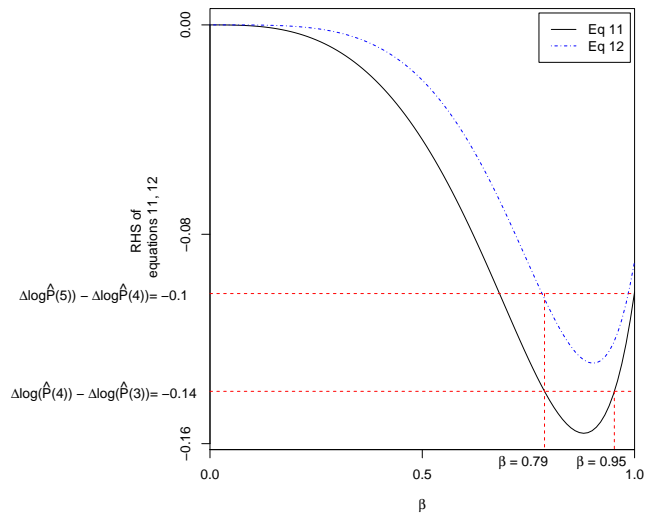


Figure 1: Plots of the right hand side of equations (11) and (12), $\beta(V(I + b) - V(I) - (V(I + b - 1) - V(I - 1)))$, at $I = 3$, and possible solutions for β . The error term is logit, and parameter values are $\nu = 0.33$, $\alpha = 1$, $M = 3$, $p = 3.31$, $b = 8$.

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 left panel of Figure 2, we plot the theoretical purchase probabilities as a function of inventory and β for the model parameterization used to construct Figure 1. The purchase probability curve flattens out as inventory increases because the disutility of a future stockout is discounted more for higher inventory values. This can be seen in the right panel of Figure 2, which plots the expected future benefit of a purchase as a function of inventory - this benefit is the value function from making a purchase, minus the value function from not purchasing, and decreases in inventory because the disutility of stocking out looms larger when inventory is lower. 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 probability function will bring the numerators and denominators of equation (7) closer together, producing an estimate of β that is closer to 1.

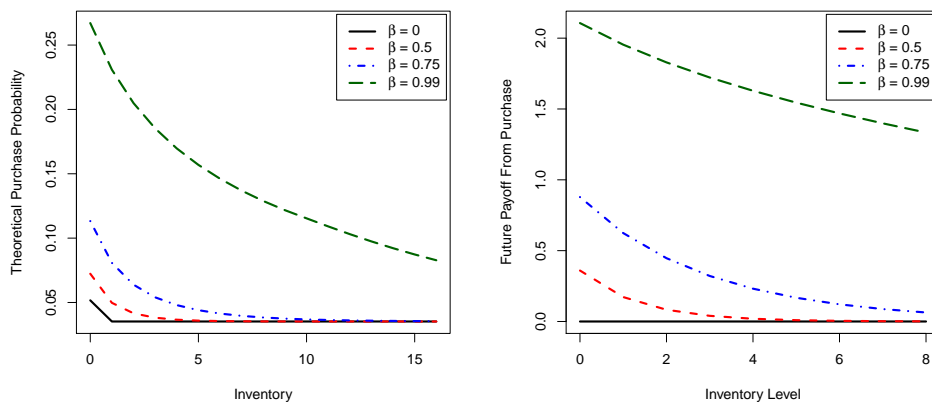


Figure 2: Theoretical purchase probability (left panel), and expected future payoff from purchase, $\beta [V(I'_1(I)) - V(I'_0(I))]$ (right panel), as a function of I and β , for storage costs of zero. The error term is logit, and parameter values $\nu = 0.4$, $M = 3$, and $p = 3.31$. Note that $I'_1(I) \equiv I - 1 + b$ 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.

5 Empirical Application

5.1 Data and Sample Construction

The empirical contributions of our work are twofold: First, we estimate consumer discount factors, along with other structural parameters, for a frequently purchased consumer packaged good. Second, we show that estimating the discount factor affects managerially relevant counterfactuals. This is an important question because past structural research on storable goods purchases has either assumed that individuals are myopic, or that they are very forward-looking. It is not known whether actual discount factors may take intermediate values between 0 and 1. To address this question, 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, and the data records purchases and shopping trips between the years 2001 through 2007, along with store-level prices. We create our estimation panel by combining IRI’s purchase panel, household trip panel, and store price panel. 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. To keep the choice set manageable, we restrict the sample to households who only purchase the top 19 brands and the top 5 sizes (50, 80, 100, 128, and 200 oz) by purchase share. If a household purchases multiple bottles within a shopping trip, we drop the household if the trip records that more than 5 bottles, or different products were purchased (these events occur very infrequently in the overall sample). We also restrict the sample to individuals 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. We do this to avoid the possibility that purchases are missing for households who do not appear to purchase for long periods of time. Our sample used for estimation contains 312 households.

After applying the exclusions described above, we compute a household level consumption rate by computing the sum of total quantity over the window for which a household is observed, and dividing by total observed weeks. We then compute inventory for each household and week of the estimation sample by assuming inventory is zero in the first week of 2001, and calculating inventory forward using the imputed consumption rates.¹² The estimation sample for the structural model corresponds to the final 3 years of data; our approach to computing initial inventory in the esti-

¹²The way we calibrate household level consumption rates and inventories is similar to Ching, Erdem, and Keane (2019). Ching, Erdem, and Keane (2019) go a step further, by taking into account the quality differences across brands, and introducing a composite measure of inventory. Their approach is particularly useful for handling households who buy multiple brands regularly.

mation sample, which makes use of the prior 4 years of data, follows the same procedure as the earlier literature on stockpiling (Erdem, Imai, and Keane (2003), Hendel and Nevo (2006)).¹³ As long as individuals do not stock out for long periods of time, which we believe is reasonable for laundry detergent, then our estimated consumption rates and imputed inventory measure should be accurate for the estimation sample. Details on the consumption rate calculation, as well as additional support for the reasonableness of our approach, are presented in Online Appendix B. There, we document that (i) our measure of imputed inventory is sensible, as it suggests that an average household holds about three bottles of detergent; (ii) decreases in inventory are highly correlated with purchase, especially if inventory is low; (iii) consumption rates are positively correlated with demographics such as household size and pet ownership; and (iv) households are not predicted to stock out very often.

In the final sample, the average number of weeks between purchases is 9.9, and the average probability of a purchase occurring is about 10%. 77% of purchases are of a single bottle, and the most popular size is the 100 oz bottle (71% of purchases), followed by the 128 oz bottle (14% of purchases), the 200 oz bottle (8.8% of purchases), and then the 50 and 80 oz sizes (less than 5% of purchases each). The first two columns of Table 1 show the purchase shares (the number of 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. Prices vary significantly across brands, but also within each brand-size combination. In the last five columns of Table 1, we show the average within-household coefficients of variation of price for all available brand-size combinations. 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. There is also a significant amount of time series price variation within a store. To show this, for each store and product in the data, we calculated the average regular store 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 around 25 percent below the regular price, and deals typically happen in 20 to 30 percent of weeks. If households are forward-looking, they will have an incentive

¹³We note that Erdem, Imai, and Keane (2003) and Hendel and Nevo (2006) assume that consumption rates may vary stochastically over time, which will imply A3 cannot be maintained. They deal with unobserved inventory by integrating out unobserved consumption using simulation, given the assumption that initial inventories are zero. In our empirical application, we do not model unobserved consumption shocks, because our numerical analysis in Appendix A suggests that the distribution of consumption shocks was difficult to separately identify from stockout costs. Additionally, our assumption of observed inventories simplifies computation of the empirical model because we do not have to simulate unobserved consumption.

Table 1: Brand Level Purchase Shares, Prices, and Price Variation

Brand	Purchase Share	Average Price (Cents Per Ounce)	Coefficient of Variation of Price (%)				
			100 oz	128 oz	200 oz	50 oz	80 oz
Tide	23.8	8.61	13.81	-	14.65	13.9	13.03
Xtra	8.9	2.48	-	10.91	7.51	-	-
Purex	9.2	4.81	17.71	15.4	11.06	10.43	-
All	7.9	5.64	12.8	-	10.64	10.92	16.58
Arm & Hammer	8.8	4.71	16.94	-	11.45	-	-
Era	5.9	5.25	17.1	-	4.8	2.46	-
Dynamo	11.7	4.58	21.1	-	14.83	-	-
Wisk	10.3	6.31	14.3	-	15.86	-	17.42
Private Label	3.8	3.51	13.89	8.76	-	5.26	-
Cheer	2.1	7.12	-	-	-	-	7.76
Fab	1.4	6.08	11.73	-	-	18.82	-
Yes	2.1	4.51	19.14	-	-	-	-
Ajax Fresh	0.4	3.19	-	12.3	-	-	-
Gain	0.5	6.16	14.06	-	7.24	-	-
Ajax	0.4	3.14	-	11.8	-	-	-
Trend	0.4	2.22	-	2.14	-	-	-
Sun	0.8	4.33	24.52	-	-	-	-
Solo	1.3	3.88	21.79	-	-	-	-
Ivory Snow	0.3	10.41	-	-	-	6.71	-

Notes: Purchase shares show the number of packages purchase of each brand in our trip-level data of 36,101 trips. Prices are computed across all household trips. For each brand-size that exists in the sample, we compute the coefficient of variation (standard deviation divided by mean) across household trips.

to stockpile in the face of the significant amount of price variation observed in the data.

5.2 Empirical Model and Estimation

The empirical model builds on the stylized model presented in Section 3 in a number of important ways. First, we allow for unobserved heterogeneity across consumers in many of the model preference parameters, including the discount factor. Second, we extend the consumer's choice set to include purchases of different brands and sizes, and allow consumers to purchase multiple bottles in a purchase occasion. Third, we model consumer expectations about the future price process and likelihood of future store visits, in addition to expectations about future inventory. Since our model

incorporates unobserved parameter heterogeneity, we use the modified Bayesian MCMC algorithm proposed by Imai, Jain, and Ching (2009) to estimate the model.¹⁴

We make one simplification of the empirical model with respect to the stylized model, which relates to the formulation of the storage cost function. There are two significant computational complications that arise in situations where individuals can choose among multiple package sizes: in our application we allow for five. The first problem is that one would need to track both the number of bottles of each size held in inventory, and model the order in which different sizes of bottles are consumed. The second problem is that if one attempted to model the storage cost function nonparametrically, the number of parameters one would have to estimate would be extremely large, leading to concerns about overparameterization. In this paper, we deal with these issues by estimating two different empirical model specifications. In our preferred model specification, presented below, we assume storage costs are zero. With this assumption, the inventory composition and the order in which bottles are consumed does not need to be modeled. We present an alternative specification in Online Appendix B, which allows for storage costs to increase with inventory. In order to keep that specification computationally tractable, we had to make a few additional simplifying assumptions related to inventory holdings. The estimated storage cost parameters from that specification suggested that storage costs are close to zero for the majority of individuals; as a result, we decide to focus on the zero storage cost specification here.

The stylized model abstracts away from a consumer’s brand choice, which can substantially increase the size of the state space, since consumers potentially have to track the holding of each brand in inventory as well as the prices of all brand-size combinations. To deal with these two issues, in our empirical model we follow Hendel and Nevo (2006) and make two simplifying assumptions: (i) consumers only care about brand differentiation at the time of purchase, and (ii) a generalized form of Inclusive Value Sufficiency proposed by Osborne (2018a), which extends Hendel and Nevo (2006), and assumes that consumers track quality-weighted averages of prices at the size level, rather than all prices at the brand-size level. 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.¹⁵ 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.

¹⁴Hendel and Nevo (2006) 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.

¹⁵Formally, 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 $\gamma_i = 0$).

We denote assumption (ii) as Generalized Inclusive Value Sufficiency (GIVS), and define it formally in Online Appendix C, Section C.1. The GIVS assumption has two key features that make it different from standard Inclusive Value Sufficiency (IVS). First, it extends IVS to incorporate multiple package purchase by assuming that the flow utility received from a particular brand scales linearly with the number of packages: The flow utility from purchasing $j = 0, \dots, J$ packages of size $x = 1, \dots, X$ and brand $k = 1, \dots, K$ is equal to $\frac{j}{J}\xi_{ikx}$, where $J = 5$, $X = 5$, $K = 19$, and ξ_{ikx} is a product-level taste coefficient. During estimation, one ξ_{ikx} coefficient is normalized to zero.¹⁶

The second key difference between GIVS and IVS relates to the distributional assumption made on the choice-specific error term. The standard formulation of IVS used in Hendel and Nevo (2006) relies on the assumption of logit errors, and assumes that consumers track the expected utility (the inclusive value) across all brands, conditional on choosing a package size. If individuals can choose to purchase multiple packages, and the error term is standard logit, then the number of packages chosen, j , enters the inclusive value in a nonlinear way. As a result, the number of inclusive values in the state space would equal the number of available package sizes, X , multiplied by the number of packages an individual can purchase, J . Instead of the logit assumption, GIVS assumes 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 j to be factored out of the expected brand-specific utility, and as a result the inclusive values can be expressed independently from j . Hence, when consumers can choose multiple packages, GIVS allows one to track only X inclusive values, rather than $J \cdot X$, which would result from standard IVS.

Under GIVS, the consumer’s flow utility function from buying $j > 0$ units of size x of brand k

¹⁶We note that for the normalized product, the mean flow utility will not scale with j , but the flow utility from the idiosyncratic error and price disutility will (see Online Appendix C.1). The assumption that utility scales linearly 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.

can be written as:

$$\begin{aligned}
& u_{it}(k, x, j, I_{it}, \varepsilon_{ijt}, p_{ikt}, c_i; \theta_i) & (13) \\
& = \begin{cases} \frac{j}{J} \xi_{ikx} - \alpha_i \cdot p_{ikxt} \cdot j + FC \cdot \mathbf{1}\{j > 0\} + \varepsilon_{ikxjt} & \text{if } I_{it} + b(x)j \geq c_i \\ \frac{j}{J} \xi_{ikx} - \nu_i \frac{c_i - (I_{it} + b(x)j)}{c_i} - \alpha_i \cdot p_{ikxt} \cdot j + FC \cdot \mathbf{1}\{j > 0\} + \varepsilon_{ikxjt} & \text{otherwise} \end{cases},
\end{aligned}$$

where the first condition is the case when a consumer has sufficient inventory to cover the consumption need c_i , and the second condition is the case when inventory is smaller than the consumption need; $b(x)$ is the number of ounces in a bottle of size x , FC is a fixed cost of purchase, and $\frac{c_i - (I_{it} + b(x)j)}{c_i}$ is the proportion of the consumption need that is not met if inventory is smaller than the consumption rate. We found it necessary to include the FC parameter in order to properly fit the low frequency of purchase we observe in the data. In the specification of flow utility for the empirical model, if there is less inventory available than c_i , we assume that the individual consumes whatever inventory is left. We multiply the stockout cost by the proportion of the consumption need that is not met to keep the model internally consistent.

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 GIVS assumption entails that an individual's expected utility over brands for choosing j packages of size x can be written as:

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

Details on the above derivation are presented in Osborne (2018a). GIVS assumes that consumers track only $\Omega_{it}(x)$, rather than each individual price p_{ikxt} .

One final addition we make to the empirical model is that we allow for the possibility that a consumer does not visit any store in some weeks, which happens in the data. We assume that store visits occur exogenously with a probability π_i , which we estimate prior to estimating the model.¹⁷ Accounting for store visits, the household's Bellman equation can be written as

¹⁷Occasionally, a household may also visit multiple stores in a week without purchasing detergent. Since our analysis is done at the weekly level, we assume that decision to not purchase is made at the store which the household most often visits.

$$\begin{aligned}
V(I_{it}, \boldsymbol{\Omega}_{it}) &= (1 - \pi_i) \left(\nu_i \frac{c_i - I_{it}}{c_i} \mathbf{1}\{I_{it} < c_i\} + \beta_i E_{\boldsymbol{\Omega}_{i,t+1} | \boldsymbol{\Omega}_{it}} V(I_{i,t+1}, \boldsymbol{\Omega}_{i,t+1}) \right) \\
&+ \pi_i \ln \left(\sum_{(j,x) \in C_1} \exp \left\{ \frac{j}{J} \Omega_{it}(x) - \nu_i \frac{c_i - (I_{it} + b(x)j)}{c_i} \mathbf{1}\{I_{it} + b(x)j < c_i\} \right. \right. \\
&\left. \left. + FC \cdot \mathbf{1}\{j > 0\} + \beta_i E_{\boldsymbol{\Omega}_{i,t+1} | \boldsymbol{\Omega}_{it}} V(I_{i,t+1}, \boldsymbol{\Omega}_{i,t+1}) \right\} \right),
\end{aligned} \tag{14}$$

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

Turning to specification details, because we include 19 brands and 5 sizes in our analysis, we separately estimate brand and size taste coefficients for many brand-size combinations.¹⁸ 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. We also allow most model parameters to be heterogeneous across the population. For the purpose of identification, we make the fixed cost of purchase, the brand and size shifters for 7 smaller share brands and 3 smaller share sizes of popular brands, homogeneous across the population (there are 39 brand-size combinations in the data). In general, in a random coefficient model, at least some coefficients must be homogeneous to guarantee identification (Ruud 1996).

The MCMC algorithm of Imai, Jain, and Ching (2009) that we employ for estimation is an iterative procedure where in each iteration one draws the model parameters from their posterior distribution, and then updates an estimate of the value function by performing a single iteration on the Bellman equation. We describe the details of the construction of the inclusive value transition process in Online Appendix C.1, of how parameters are drawn in Online Appendix C.2, and of the value function updating in Online Appendix C.3.

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 (exogenous consumption rates) and X2 (discontinuous storage costs) will generally be informed by theory or institutional knowledge. We believe X1 is behaviorally sensible because bottles of laundry

¹⁸To be precise, we assume that $\xi_{ikx} = \xi_{ik} + \xi_{ix}$ for most products. We made this simplification because the number of purchases for many brand-size combinations was small, which made it difficult to identify unrestricted brand-size coefficients for these products. We provide exact details on where this restriction is relaxed in the discussion in Online Appendix D.

detergent come with guidelines on how much to use when washing a load, and so individuals are unlikely to gain or lose additional utility by deviating from these guidelines. X2 also applies to our setting, since liquid laundry detergents are sold in plastic bottles of fixed sizes. Turning to X3 (at least 3 periods are necessary to use up a package), our implied consumption rates suggest that the majority of households do at most 4 or 5 loads of laundry per week. The most popular sizes of detergent contain 32 or 64 loads, and so it will take an average household 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. Details on consumption rate levels are provided in Online Appendix B.

Identification of the stockpiling model also requires that Rank condition R1 holds. This rank condition essentially requires that the probability of purchase as a function of inventory becomes more steeply sloped as inventory decreases. We provide support for this with a reduced-form exercise where we run a regression of a dummy variable for purchase on a flexible function of inventory. The function of inventory we use allows the purchase probability to be different depending on whether inventory is 0, 0 to 50 ounces, 50 to 100 ounces, 100 to 200 ounces, or more than 200 ounces. The estimated purchase probability increases at an increasing rate as inventory declines, and sharply increases when inventory hits 0. This pattern is qualitatively consistent with the plots in Figure 2 and consistent with substantial stockout costs (Assumption X1). The exact specification, along with parameter estimates, are presented in Online Appendix B, Table B.1.

A final comment relates to the identification of unobserved parameter heterogeneity: Our identification arguments still apply if we only observe data for one individual, provided that the number of observations is sufficiently large. Here, we estimate our model on 3 years of weekly scanner panel data. Hence, we allow for heterogeneous parameters in the estimation. To provide more evidence that the heterogeneity can be identified, we perform an artificial data experiment in Online Appendix E, where we simulate purchases given our estimated parameters from Section 5.4, and show our estimation procedure recovers the population distribution of the utility parameters quite well.

5.4 Estimation Results

To obtain parameter estimates, we run our MCMC sampler for 20,000 iterations, and save the final 10,000 draws. The MCMC algorithm appears to converge after about 2,000 to 3,000 draws, so our cutoff point is conservative. In Table 2, we present estimates of the moments of the population distribution of the parameters that affect an individual’s dynamic decision: the price coefficient, stockout cost, discount factor, and fixed cost of purchase. The table presents the 33rd, 50th, and 66th percentiles as well as the population mean, and shows there is a significant amount of

Table 2: Dynamic Parameter Estimates

Parameter	1st Tertile	Median	Mean	2nd Tertile
Price Coefficient	-0.29 [-0.31, -0.28]	-0.24 [-0.26, -0.23]	-0.27 [-0.29, -0.26]	-0.21 [-0.22, -0.2]
Stockout Cost	0.29 [0.24, 0.36]	0.39 [0.31, 0.49]	0.48 [0.37, 0.66]	0.5 [0.4, 0.67]
Discount Factor	0.62 [0.14, 0.89]	0.94 [0.81, 0.98]	0.71 [0.58, 0.82]	0.99 [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 column 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.

individual-level heterogeneity in all of the estimated 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. Estimates of the product-specific taste coefficients are shown in Table D.6 of Online Appendix D.

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 is around 0.99, while 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 3, 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. 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 also estimated an auxiliary specification where we included demographic interactions, but found that

the demographic interactions did not significantly affect the discount factor.

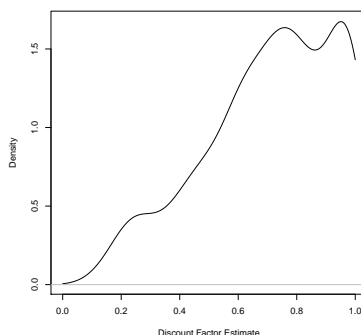


Figure 3: Kernel Density of Individual-Specific Discount Factor Estimates.

Our findings that estimated discount factors are lower than the calibrated benchmark, and that there is a significant amount of heterogeneity across individuals, are both consistent with recent work that has estimated discount factors. In the context of cell phone usage, Yao, Mela, Chiang, and Chen (2012) find that the weekly discount factor is 0.91, and in one of their robustness exercises they find that the population distribution of discount factors has a standard deviation of 0.26. For Blu-Ray player purchases, Dubé, Hitsch, and Jindal (2014) find average annual discount factors of about 0.4 (which is around 0.97 after converting to a weekly discount factor), and a significant amount of heterogeneity across individuals. Moreover, Yang and Ching (2014) document (in their Appendix B) that Italian consumers’ annual discount factors range from 0.8 to 1 when considering intertemporal trade-offs of one’s annual income. Taken together, this pattern suggests that as the stakes involved in a dynamic decision increase, consumers may be willing to spend more mental resources to plan further ahead, which will be reflected in a higher discount factor. We should note that a discount factor in the range of our estimated value is consistent with a setting where consumers plan several weeks ahead when they make their purchase decisions. This seems reasonable for laundry detergent. In contrast to what the standard calibrated discount factor (i.e., a weekly discount factor 0.9995) would imply, it seems implausible that consumers plan years ahead for laundry detergent purchases.

Earlier research on dynamic discrete choice models has either assumed that individuals are very forward-looking, or completely myopic. To gauge these alternative assumptions, we estimate two additional specifications: (i) $\beta = 0.9995$ for all households; here we calibrate β using an annual interest rate of 5%; (ii) a myopic specification where $\beta = 0$ for all households. In both specifications, we estimate all parameters other than β . Estimates of the dynamic parameters for each of these specifications are shown in Table 3, along with the marginal log-likelihood and

Deviance Information Criterion (DIC). Both of these models provide a worse fit to the data, in terms of both the marginal log-likelihood and DIC, relative to the specification where the discount factor is estimated. In terms of parameter estimates, most are similar across the specifications, with the exception of the stockout cost. The $\beta = 0.9995$ model produces lower estimated stockout costs: If individuals discount the future at a higher rate, the stockout cost will have a bigger effect on purchase likelihoods. Consistent with this reasoning, the myopic model produces significantly higher estimates of stockout costs.

The model specification with estimated discount factors also provides a better fit to many moments of the data, relative to the two more restricted specifications. In particular, we find that the main specification provides an improved fit over the $\beta = 0.9995$ specification to interpurchase times, purchase probabilities given inventory levels, and price sensitivity. When compared to the myopic specification, the main specification produces substantially more accurate predictions of product level market shares, and does a better job of capturing how purchase probabilities vary with inventory. More detailed comparisons of model fits are provided in Online Appendix F.

In addition to the model specifications discussed above, we perform a number of additional estimation exercises to verify that our estimate of low discount factors is robust to changes in the consumption rate. In particular, we re-estimate the model with consumption rates set to 5%, 15%, and 25% higher and lower than the estimated rates. For differences of 5%, the estimated model parameters look very similar to the main specification. There are larger differences in the estimated parameters for bigger changes to the consumption rates, but this is not surprising because increases or decreases of 15% or more imply either unrealistically large inventory holdings, or unrealistically long periods of stocking out. Nonetheless, even in these cases our qualitative findings of low discount factors is maintained, and the mean estimates almost always lie within the 95% confidence bounds of the main model. We also note that the model specification with increasing storage costs produces lower discount factor estimates (with a population average of 0.55). Notably, that version of the model provides a worse fit to the data, likely due to the additional restrictions related to inventory bounds and size choice. A full description of these robustness exercises is discussed in Online Appendix B.

6 Counterfactuals

There is a long history of research in empirical marketing that quantifies the impact of changing the time series process of promotional prices on product sales (Mela, Jedidi, and Bowman (1998), Jedidi, Mela, and Gupta (1999), Liu and Balachander (2014), Osborne (2018b)). Correctly forecasting the

Table 3: Dynamic Parameter Estimates: Alternative Assumptions on The Discount Factor

Discount Factor Fixed at 0.9995 (Interest Rate Calibrated)				
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			
Discount Factor Fixed at 0 (Myopic)				
Price Coefficient	-0.28	-0.23	-0.26	-0.2
	[-0.29, -0.27]	[-0.24, -0.22]	[-0.27, -0.25]	[-0.21, -0.19]
Stockout Cost	0.74	0.9	1.01	1.09
	[0.64, 0.83]	[0.8, 0.99]	[0.91, 1.1]	[0.98, 1.19]
Fixed Cost of Purchase	-	-	-1.77	-
			[-1.85, -1.69]	
Log-likelihood	-19629.16			
Deviance Information Criterion	40560.51			

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 column 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.

impact of such policy changes on product sales is important to retail managers. Previous structural work has made these forecasts under the assumption that β is set to its calibrated value of 0.9995. In this section, we will quantify how much bias this standard approach could entail if the true value of β is different from its calibrated value.

To quantify this bias, we will compute the impact of two counterfactual changes to the price process on product sales, for the model specification with estimated discount factors, as well as the specification with $\beta = 0.9995$. The first counterfactual price change is to increase the depth of promotions, and the second is to increase the frequency of promotions. Moreover, because consumer price expectations may take time to adjust, we will perform the comparison under two different scenarios: (i) a short-term effect, where price expectations have not adjusted yet, and remain at

Table 4: Counterfactual Effect of Increased Promotional Depth/Frequency on Quantities and Revenues

Increased Promotional Depth				
	Estimated Discount Factor		$\beta = 0.9995$	
Counterfactual	Quantity	Revenue	Quantity	Revenue
Long-Term	502.72	1018.48	576.95	1183.98
	[448, 560]	[777.49, 1256.67]	[514, 641]	[916.38, 1452.15]
Short-Term	507.56	1034.54	599.05	1255.78
	[452, 566]	[794.98, 1269.93]	[532, 669]	[972.66, 1542.56]
Increased Promotional Frequency				
	Estimated Discount Factor		$\beta = 0.9995$	
Counterfactual	Quantity	Revenue	Quantity	Revenue
Long-Term	49.34	232.99	55.19	276.34
	[-11, 110]	[-214.35, 679.05]	[-7, 118]	[-171.78, 734.91]
Short-Term	49.62	234.77	55.83	279.28
	[-11, 111]	[-213.24, 681.19]	[-7, 119]	[-172.65, 735.99]

Notes: This table shows the simulated change in total number of units sold (quantity) and revenues for the 100 oz bottle of Tide over the 3 year period of the estimation sample. Changes are shown for each of the counterfactual price processes in the panel headings, compared to simulated choices from the prices observed in the data. In the row ‘Long-Term’, consumers’ beliefs about the price process reflect the counterfactual process. In the row ‘Short-Term’, consumers’ beliefs about the price process correspond to those in the original data, rather than the counterfactual prices. 95% confidence bounds are shown in brackets.

the actual observed price process; (ii) a long-term effect, where consumer price expectations have adjusted to the counterfactual process. We compute both short-term and long-term effects because often times, firms need to evaluate a new price promotion policy soon after its implementation. In a short time interval, consumer expectations may not have adjusted to the new price process.

We compute counterfactual price changes for the 100 ounce Tide bottle. To increase promotional depth, we identify all promotional prices for the focal UPC, and then divide the prices by 2. To increase promotional frequency, our procedure is to first identify deals in the store data using IRI’s supplied price reduction flag. We double the number of store-upc date observations where a deal is observed by randomly assigning some non-deal observations to be deals. In particular, if there are N deal observations in the original price data, we leave alone those N deals, and randomly assign another N non-deal observations to be deals, so the total number of deals in the counterfactual

price series is $2N$. 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. When we compute short-term counterfactuals, we update the observed price process, but assume that consumer expectations about future prices do not change, and correspond to those in the original data. In the long-term counterfactuals, consumer expectations are fully adjusted and correspond to the counterfactual price process.

The simulated changes in quantities and revenues for 100 oz Tide are shown in Table 4.¹⁹ The top panel shows the changes for the increased depth counterfactual, and the bottom for increased frequency. The first two columns show the effects when discount factors are estimated. When promotions are deepened, the main specification suggests that in the long-term the number of bottles sold increases by 502 units; the short-term effect (shown in the second row) is slightly larger than the long-term effect. In the short-term, individuals do not realize that deeper promotions will occur on a regular basis, and stockpile even more in response to the observed deal. The counterfactual price change does not lead to a large increase in quantity sold for the entire laundry detergent category - on the order of 100 units. Hence, the increase in Tide’s quantity comes at the expense of sales for competing products. We note that there are two mechanisms by which forward-looking behavior can result in consumers switching to Tide. The first is that if Tide increases its promotional depth, when individuals purchase on sale they will purchase more units and will be less likely to purchase other brands in the future, due to having higher inventories. The second effect is driven by price expectations: if a forward-looking individual who is running out of detergent observes a price promotion on a brand that competes with Tide, she may skip the promotion and wait for Tide’s promotion, because she knows the Tide promotion will be better. Increasing promotional frequency will also increase quantity purchased and revenues, although the confidence bounds of the effect of this counterfactual change contain zero.

The results of this counterfactual analysis shed light on the behavioral mechanism underlying the effects of promotions identified by Jedidi, Mela, and Gupta (1999). Jedidi, Mela, and Gupta (1999) find that improved promotions for a brand make it less necessary for a consumer to switch brands,

¹⁹We compute counterfactual quantities and revenues given a counterfactual price series by simulating choices at each parameter draw, allowing us to put confidence bounds around the counterfactuals. The computational burden of the counterfactual exercise arises from having to compute individual-specific value functions at each parameter draw. To mitigate this burden, we follow a procedure outlined in Ching, Imai, Ishihara, and Jain (2018), where they propose that one should iterate through the saved parameter draws from the MCMC sampler. At each saved draw, we update the value function once at the counterfactual prices given an approximate value function constructed from saved value functions, and simulate choices. The value functions computed in this way will converge to the true value functions after the saved draws have converged. We show the counterfactuals for the last 10,000 draws.

and argue that the behavioral mechanism behind this finding is that forward-looking consumers will be more likely to skip promotions on competing brands, and will instead wait for a better promotion on the focal brand. As we describe above, our structural stockpiling model explicitly models this mechanism.

The second two columns of Table 4 show the counterfactual results when we fix $\beta = 0.9995$ and estimate the rest of the structural parameters. In this standard approach, one would substantially overpredict the increase in units sold, by 18% in the short-term, and 15% in the long-term. This overprediction mainly occurs because more forward-looking consumers are also more responsive to the improved price process. The extent of bias in the forecasted quantity sold is larger in the short-term. This is because before consumers adjust their price expectation, they think a deep discount is highly unlikely. Therefore, in the short-term, consumers with higher values of β will have a stronger incentive to take advantage of a better discount and stockpile more. The model with $\beta = 0.9995$ also overpredicts the effectiveness of increased promotional frequency, but by a smaller amount (about 10%). Compared the $\beta = 0.9995$ specification, in the specification where the discount factor is estimated, the differences between the predicted short-term and long-term effects are small. In that specification, our estimate of the discount factor is low, which may explain why expectations have little impact on the counterfactual predictions.

Table 5 further demonstrates how one’s responsiveness to promotional depth varies with the discount factor. In this table, we present the long-term predicted change in quantity (number of bottles of 100 oz Tide purchased by a consumer), purchase frequency (overall number of predicted purchases divided by number of weeks), and revenues (for 100 oz Tide) per consumer, for individuals whose estimated discount factor is above the median of the individual-specific estimates (first row) or below the median (second row).²⁰ The results are intuitive. In response to an increase in promotional depth, quantity purchased and revenue increases significantly, while the change in purchase frequency is positive but very small. Moreover, individuals with higher discount factors buy more. The results in Table 5 show that most of the increase in consumer demand documented in Table 4 is driven by the behavior of forward-looking consumers. These results also provide additional support for the conjecture of Jedidi, Mela, and Gupta (1999), which relies on forward-looking behavior.

²⁰In the short-term, we find similar patterns but slightly stronger effects, especially for more forward-looking consumers.

Table 5: Response to Increase in Promotional Depth Given Discount Factor Estimate

Change in	Tide	Purchase	Tide
	Quantity	Frequency	Revenues
Discount Factor			
Above Median	1.7106	0.0024	2.9849
Discount Factor			
Below Median	1.5116	0.0034	1.1879

Notes: The first row of the table shows the average simulated change in Tide’s quantity (number of 100 oz bottles), purchase frequency (fraction of weeks where a purchase occurs), and revenues, for individuals whose estimated discount factor is above the median of the individual-specific values. The second shows the same predicted changes for individuals whose estimated discount factors are below the median of the individual-specific values. The number of observations is 312.

7 Conclusion and Discussion

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 (2006)) has imposed smoothness assumptions on the storage cost function, assuming it to be continuous in inventory. We emphasize that this seemingly innocuous simplifying assumption on storage cost rules out exclusion restrictions which naturally arise from the institutional features of this problem. This is why previous work has assumed that consumers are homogeneous in their discount factors, and the discount factor is set according to the prevailing interest rate. By properly modeling 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. Our finding that consumer discount factors are heterogeneous and much lower than the interest-rate calibrated benchmark of 0.9995 has important implications for future research. In particular, it suggests that if possible, researchers should estimate, rather than calibrate discount factors.

The preceding discussion suggests at least two possible avenues for future research. One is that in order to understand whether our finding related to estimated discount factors generalizes to other categories, one could identify additional product categories where the discount factor could

be estimated. The results of this exercise would shed light on the extent to which discount factors are context-dependent. This research direction could also improve our understanding about what drives one's forward-looking behavior. An additional avenue for research stems from the fact that our proof is for the case of observed inventory, which we feel is a reasonable assumption for the laundry detergent category. The set of categories where such analysis could be performed could be broadened if a formal proof of identification could be developed for situations where inventory is unobserved, due to stochastic consumption shocks or endogenous consumption. In Appendix A, we provide evidence from numerical analysis of an extension to the model presented in Section 3 that identification may be possible for this case, but leave a formal proof to future research.

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., T. Erdem, and M. P. Keane (2019). How much do consumers know about the quality of products? evidence from the diaper market. Forthcoming in Japanese Economic Review.
- Ching, A. and M. Ishihara (2018). Identification of dynamic models of rewards program. Japanese Economic Review 69(3), 306–323.
- Ching, A. T., 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. T., S. Imai, M. Ishihara, and N. Jain (2018). A practitioner’s guide to bayesian estimation of discrete choice dynamic programming models: Counterfactual policy experiments. Working Paper, Johns Hopkins University.
- 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.
- Dalton, C. M., G. Gowrisankaran, and R. Town (2019). Salience, myopia, and complex dynamic incentives: Evidence from medicare part d. Forthcoming in the Review of Economic Studies.
- De Groote, O. and F. Verboven (2019). Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems. American Economic Review 109(6), 2137–72.
- Dubé, J.-P., G. Hitsch, and P. Jindal (2014). 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.
- Fang, H. and Y. Wang (2015, May). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. International Economic Review 56(2), 565–596.
- Frederick, S., G. Loewenstein, and T. O’Donoghue (2002). Time discounting and time preference: A critical review. Journal of Economic 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*.
- Haviv, A. (2014). Does purchase without search explain counter cyclic pricing? working paper, Simon Business School, University of Rochester.
- Hendel, I. and A. Nevo (2006). Measuring the implications of sales and consumer inventory behavior. Econometrica 74, 1637–1673.
- Hendel, I. and A. Nevo (2013). Intertemporal price discrimination in storable goods markets. American Economic Review 103(7), 2722–2751.

- 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>.
- Jedidi, K., C. Mela, and S. Gupta (1999). Managing advertising and promotion for long-run profitability. Marketing Science 18(1), 1–22.
- 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.
- Mela, C. F., K. Jedidi, and D. Bowman (1998). The long-term impact of promotions on consumer stockpiling behavior. Journal of Marketing Research 35(2), 250–262.
- 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. (2018, September). Lower price or higher reward? measuring the effect of consumers’ preferences on reward programs. Management Science 64(9), 4451–4470.

- 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.
- Sun, B. (2005). Promotion effect on endogenous consumption. Marketing Science 24, 430–443.
- Wang, E. Y. (2015). The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. The RAND Journal of Economics 46(2), 409–441.
- 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.

A Appendix: Identification with Unobserved Inventory

Prior empirical work in stockpiling has assumed that consumption needs vary stochastically, and are unobserved to the researcher (see Erdem, Imai, and Keane (2003) or Hendel and Nevo (2006)), which implies a failure of Assumption A3. In this section, we provide an informal discussion of how identification may still be obtained if consumption is exogenous, but stochastic. We note that Hendel and Nevo (2006) assume endogenous consumption (an individual's optimal consumption is a function of inventory); we leave this case to future research.

In the case of observed inventory, the moments that identify the model parameters are the choice probabilities at different levels of inventory. In the case of unobserved inventory, the analog to these moments are the choice probabilities given the amount of time elapsed since the last purchase occurred, which is the purchase hazard. Because constructing the purchase hazard involves integrating out unobserved consumption shocks and inventories held at the time of the last purchase, there is no straightforward way to compute choice-specific values from moments and solve for the model parameters directly. As a result, we discuss intuitively how the properties of the purchase hazard may help identify the model parameters, based on numerical solutions of an extended version of the model presented in Section 3. In the extended model, we set $M = 3$, $b = 8$, and assume the consumption need, c , is stochastic, i.i.d. across time and individuals, and takes on the value 1 with probability π_c and 2 with probability $1 - \pi_c$.

To aid the discussion of identification, we present the simulated purchase hazard for a similar model parameterization as in Figure 2, but with stochastic consumption rates where $c \in \{1, 2\}$ and $\pi_c = 0.5$. To better illustrate the effect of changing model parameters on the purchase hazard, we lower the stockout cost to 0.25 from 0.4, and the price to 2 from 3.31. When we compute the purchase hazard, we integrate over unobserved inventory at the time of purchase through simulation, where inventory is drawn from the steady state distribution of inventory in the population. We also integrate out consumption shocks that occur after purchase. The plots in Figure 4 show the purchase probabilities averaged over both unobserved inventories and unobserved consumption shock draws. It is important to note that the population distribution of inventory at the time of purchase ranges from 0 to 2 full packages. In particular, some consumers may be very close to using up one package.

We can provide a straightforward intuitive argument for the identification of β if the researcher is willing to restrict the storage cost parameters to be zero (i.e., $\omega_1 = \omega_2 = \omega_3 = 0$). The purchase hazard for this case is presented in the left plot of Figure 4. Focusing first on the purchase hazard when $\beta = 0$, it can be seen that the purchase probability does not change for the first three periods after a purchase. In contrast, for forward-looking individuals the purchase hazard increases during

the first three periods. The intuition behind this difference is that in the first few periods after a purchase occurs in period 0, flow utility does not change with inventory, and so when $\beta = 0$ the purchase probability does not change; raising the discount factor increases the future value of a purchase and increases the purchase hazard's slope during these periods. Hence, if storage costs are zero, the slope of the purchase hazard in the first few periods provides identifying information about the discount factor. It is also notable that when $\beta = 0$, the purchase hazard is s-shaped, and raising β smooths out the purchase hazard. Intuitively, a myopic consumer is not willing to trade off future utility for current utility, so her purchase hazard will start to rise sharply after 4 periods when individuals start facing stock outs. In particular, individuals who have no inventory at the time of purchase (period 0), and receive four consecutive draws of $c = 2$ in a row, will stock out in period 4, since $b = 8$. Hence, the purchase probability will start to increase after this point. 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. The extent to which the purchase hazard is s-shaped will also provide identifying information about β .

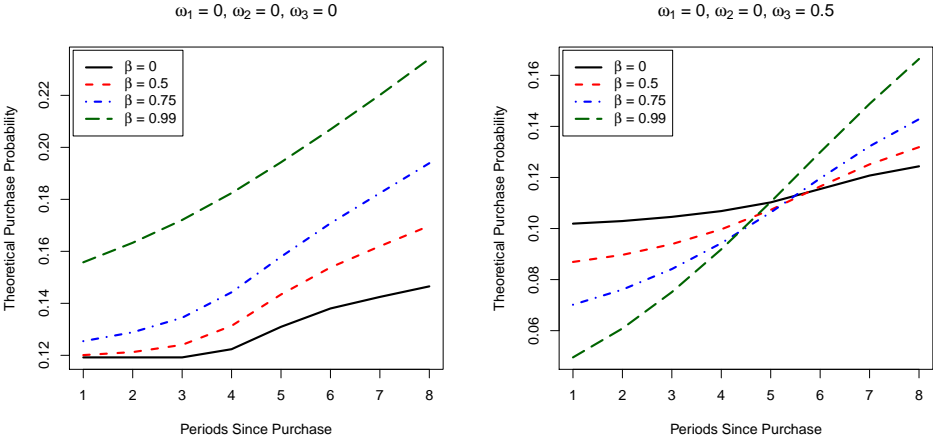


Figure 4: Probability of purchase as a function of the number of periods since the last purchase occurred (period 0). Parameter values $\nu = 0.25, \pi_c = 0.5, M = 3, p = 2$, and logit error term, for different values of the discount factor and storage cost.

Turning to the identification of the other model parameters, ν, π_c , and α , we also investigated the impact of changing both ν and π_c on the purchase hazard, and found that for low values of β , these parameters had very similar effects. Decreasing π_c will increase the rate at which an individual runs out, and shift up the purchase hazard in later periods, while having little effect on it in earlier periods. Increasing ν also shifts up the purchase hazard in later periods, but has little effect in earlier periods, because increasing this parameter makes stockouts more costly. As a result,

we suspect these two parameters may be difficult to separately identify. We note that Hendel and Nevo (2006) normalize the distribution of consumption shocks in their empirical application, likely as a result of such an identification problem. In practice, researchers may wish to normalize one of these parameters; the other parameter can then be identified by the level of the purchase hazard in later periods. Finally, increasing or decreasing α will shift the purchase hazard up and down for all periods, because α affects the overall purchase probability. In practice, price variation would help identify α .

We have not found a straightforward argument for the identification of β when the researcher wishes to estimate the storage cost parameters in addition to β , because increasing one of the ω_B parameters and increasing β can have similar effects on the purchase hazard: Increasing either parameter will increase both the purchase hazard's slope in periods soon after purchase, and make it smoother. To see this, consider the right plot of Figure 4, which shows how the purchase hazard is affected by changing β when we set the storage cost for three packages to be positive (i.e., $\omega_3 > 0$), but we keep $\omega_1 = \omega_2 = 0$. Comparing the left and right plots, for a given value of β , the purchase hazard's slope in the first few periods is higher when ω_3 is positive. Increasing the storage cost will increase the slope of the purchase hazard, even for a myopic individual, because inventory is unobserved to the researcher at the time of purchase (period 0). Recall that in the left plot, the increase in purchase hazard at period 4 when $\beta = 0$ is driven by individuals with no inventory at the time of purchase. In the right plot, the initial increase in the purchase hazard is driven by individuals who have 3 packages in inventory, but are close to using up one of them. Some of these individuals will use up a package shortly after period 0, leading to a decrease in storage cost from 0.5 to 0, and a corresponding increase in the purchase probability. We also note that when storage costs are higher, the purchase hazard appears smoother for each level of β .

When inventory is unobserved, the exclusion restrictions can help fulfill a necessary condition for identification, which is that the number of periods where the purchase hazard is observed must be greater than the number of parameters being estimated. The exclusion restrictions will limit the number of storage cost parameters to M , and as long as the purchase hazard can be computed for at least $M + 3$ time periods, this necessary condition will be satisfied. Assuming without loss of generality that π_c is normalized, the model parameters to be estimated will be the storage costs, $\omega_1, \dots, \omega_M, \alpha, \beta$, and ν . A sufficient condition for identification of β would be that one could not completely fit the purchase hazard with the M storage cost parameters, ν , and α . The variation in the purchase hazard that is not explained by storage costs, ν , and α will inform the size of the discount factor.