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

Andrew Ching* and Matthew Osborne†

July 13, 2015

Abstract

We develop a new empirical strategy for identifying the parameters of dynamic structural models in markets for storable goods, with a focus on identification of the discount factor. The identification strategy rests on an exclusion restriction generated by discontinuities in package sizes: In storable goods product categories where consumption rates are exogenous and package sizes are discrete, current utility does not depend on inventory unless a package gets used up. Most of the time, inventory only enters a consumer’s expected future value. We develop conditions for identification in two situations: when inventory is observed to the researcher, and when it is unobserved. When inventory is unobserved, we show that the shape and slope of the purchase hazard can identify the discount factor. We demonstrate the feasibility of our identification strategy with an empirical exercise, where we estimate a stockpiling model using scanner data on laundry detergents. Preliminary estimates suggest that consumers are not as forward-looking as most papers in the literature assumes; our estimates of weekly discount factors average at about 0.91, which is significantly lower than the value used in previous research (it typically is set at 0.99, using the market interest rate). We also find evidence of significant unobserved heterogeneity in discount factors, estimating the 25th and 75th percentiles of the population distribution of discount factors to be 0.88 and 0.96.

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

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

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

Forward-looking behavior is a critical component of many quantitative models of consumer behavior used by researchers in marketing and economics (Erdem and Keane (1996), Crawford and Shum (2005), Hendel and Nevo (2006a), Erdem, Imai, and Keane (2003), Seiler (2013) and Liu and Balachander (2014), Osborne (2011), Yang and Ching (2014)). When consumers are forward-looking, they also behave strategically when making their purchase decisions. For instance, in the context of new durable goods such as cameras or smart phones, consumers may wait to purchase a product if they expect the product’s price to fall in the future. Similarly, in the context of storable packaged goods such as canned tuna or canned soup, forward-looking consumers may respond to a temporary price promotion today by stockpiling the product, since they understand that future prices are likely to be high (e.g., Haviv (2014)). If all shoppers are extremely forward-looking and act in such a saavy way, durable goods producers would not be able to use a price skimming strategy (Coase 1972), and grocery stores or supermarkets would never sell their carried items at a regular price. This is not the case, since in reality price skimming and hi-lo pricing are both prevalent, and consumers do make purchases of products when their prices are high. Prior research on periodic promotions in economics (Hendel and Nevo (2013), Hong, McAfee, and Nayyar (2002), Pesendorfer (2002), Sobel (1984)) has recognized that firms can use periodic promotions to price discriminate between patient and impatient consumers. Thus, the extent to which consumers consider the future clearly has implications for firms’ optimal pricing strategies. Additionally, forward-looking behavior has important implications for public policy. One such area is in estimating price indexes, such as the Consumer Price Index. Price indexes are constructed by government agencies to measure inflation, and are used by businesses to index contracts. Standard price indexes will correctly measure changes in the cost of living if consumers are myopic; however, if consumers substitute purchases across time, recent research has suggested that standard indexes overstate growth in the cost of living (Feenstra and Shapiro (2003), Reis (2009), Osborne (2014)).

In these models, the strength of forward-looking behavior is captured by a parameter called a discount factor: the closer the discount factor is to 1, the more weight consumers put on future payoffs when making current decisions. However, instead of estimating the discount factor, much research focuses on estimated models of dynamic consumer behavior exercises the “rational expectations” assumption and uses the prevailing interest rate to fix the discount factor accordingly. Depending on the length of a period, this calibration approach would lead to a value of weekly discount factor about 0.96-0.99.¹ Interestingly, Frederick, Loewenstein, and O’Donoghue

¹ At a yearly interest rate of 5%, a rational consumer would discount utility in the following year at a rate of about
(2002) survey prior experimental work in measuring discount factors: They document a significant amount of heterogeneity in the estimates those studies obtained, ranging between close to 0 and close to 1. Additionally, in stated choice experiments performed by Dubé, Hitsch, and Jindal (2014), consumers appear to be much less forward-looking than economic theory implies, with average discount rates of 0.43. Dubé, Hitsch, and Jindal (2014) also find substantial heterogeneity in discount factors across individuals.

The reason why the discount factor is typically not estimated in structural econometric work stems from an identification problem: Most problems under study do not provide natural exclusion restrictions that could help identify this parameter, and so any estimate of the discount factor would be heavily reliant on functional form assumptions. Roughly speaking, to address this problem one would need to have at least one state variable that impacts a consumer’s future payoffs, but not her current payoffs. In econometrics terminology, such a variable provides exclusion restrictions that helps to identify the discount factor (since it is excluded from current payoffs but not future payoffs). The intuition is that if a consumer is completely myopic, then the consumer’s choice should be independent of that variable. The extent to which consumer’s choice is influenced by the exclusion restriction provides information about how forward-looking consumer is. To our knowledge, there are only a handful papers that explore such an identification argument to estimate consumer’s discount factor or her incentive to consider future payoffs (Ishihara and Ching (2012), Chung, Steenburgh, and Sudhir (2013), Lee (2013), Fang and Wang (2015), Ching, Imai, Ishihara, and Jain (2012), Ching, Erdem, and Keane (2014), Chevalier and Goolsbee (2009), Magnac and Thesmar (2002)).

One dynamic consumer choice problem that has received much attention recently is consumer stockpiling of storable goods. However, almost all papers in this literature calibrate the discount factor using interest rates instead of estimating it using choice data (i.e., revealed preference). We contribute to this literature by arguing that one of the key state variables of this problem, consumer inventory, provides exclusion restrictions that can help identify the discount factor. We illustrate how using an example drawn from the laundry detergent market. Suppose a consumer is down to her last bottle of laundry detergent. She washes one full load of clothes per week (and such need

0.95, and would have a weekly discount rate of about 0.999.

2 Complementary work by Akça and Otter (2015) describes an alternative mechanism by which inventory can be used to identify the discount factor: the order in which brands are consumed can be used for identification. Our approach focuses on package size discontinuities, rather than the order in which brands are purchased. Geweke and Keane (2000) and Yao, Mela, Chiang, and Chen (2012) explore another identification strategy which requires making assumptions that the current payoffs are either observed or can be recovered from a static environment first.
is driven by her habit of wearing clean clothes every day). 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 becomes stronger as inventory (i.e., the amount of detergent in the bottle) runs down,
and her demand would appear to become more sensitive to price cuts. Moreover, for any amount
of inventory remaining, the more forward-looking a consumer is, the more intense this feeling of
urgency will get.

However, note that if a consumer is totally myopic, then inventory should not affect her behavior,
unless she runs out. A myopic consumer will only care about having enough detergent to do the
current week’s laundry, and her current utility will be only affected by the storage cost, which does not change since she still has a single bottle taking up the same amount of space. The example illustrates that inventory can provide exclusion restrictions to help identify the consumer’s discount factor, because inventory impacts the consumer’s expected future payoff, but not her current payoff. Importantly, even though the inventory goes down as she keeps consuming, the storage costs do not change because she still needs to keep this last bottle and the space it takes up does not shrink with the inventory in general.

It is important to highlight that prior work which estimates structural stockpiling models has
failed to recognize that storage cost is typically not a continuous function of inventory, and rather
assumes that storage cost is a linear or quadratic function of quantity held. This distinction is
critical, because if storage cost is a continuous function of inventory, then it is not clear that the
theoretical prediction we outlined in the previous two paragraphs would hold. This is because a
continuous storage cost gives a consumer an incentive to wait longer before buying a new bottle,
since the storage cost keeps dropping as the inventory shrinks.

We formally demonstrate the above intuition in a stylized setting by proving that a consumer’s expected future value of purchase decreases in inventory, and this expected future value rises as her discount factor rises. This will imply that a forward-looking consumer will be more likely to make a purchase before she runs out than a myopic consumer. A forward-looking consumer will also become more responsive to price cuts as her inventory drops, while a myopic consumer’s sensitivity to price cuts should remain constant until she runs out. We then derive conditions under which, if inventory were observed to the researcher, one could identify the discount factor. The formal proof relies is similar to that proposed by Fang and Wang (2015), and relies on the ability of the researcher to compute choice probabilities at different levels of inventory. Intuitively, if inventory is observed, the researcher can compute the probability a consumer makes a purchase at each level of inventory. For a forward-looking consumer, this probability should rise smoothly as her inventory
drops; for a myopic consumer, it will not change as inventory drops, until she runs out.

Much research that estimates structural stockpiling models uses supermarket scanner data, which does not track consumer inventory. Hence, a key issue is that the main state variable of interest, inventory, is unobserved to the researcher. Formal proofs of identification of the discount factor that use exclusion restrictions (Fang and Wang (2015), Magnac and Thesmar (2002)) rely on the researcher being able to observe the same state variables that the consumer does. Since the standard identification proof does not apply when inventory is unobserved, we will argue that a different set of moments than choice probabilities at observed states can still lead to identification of the discount factor. To see how our approach works, first note that when inventory is observed, the change in the purchase probability that is observed as inventory decreases can help identify the discount factor. We argue that when inventory is unobserved, a natural analog to this probability is the purchase hazard, which is defined to be the average probability of a purchase occurring \( \tau \) periods after a purchase occurred in period \( t \), with no purchase occurring in the intervening time.

To see the intuition for this, note that we know that inventory increased in period \( t \), and on average it must decrease over time. In the absence of storage costs, if consumers are myopic the purchase hazard will be flat until consumers begin to run out. If consumers are forward-looking then the purchase hazard will rise smoothly over time. We introduce a set of assumptions necessary for formal identification in the presence of storage costs, and argue that having exclusion restrictions is still an important sufficient condition for identification. We demonstrate that the discount factor is still identified when inventory is unobserved and storage costs are present in two ways: First, by showing that one can only properly fit the slope and the curvature of the observed purchase hazard with the correct discount factor. Intuitively, as consumers get more forward-looking, the slope of the purchase hazard rises in the first few periods after a purchase, and the curvature of the purchase hazard becomes smoother since consumers are more willing to trade off future and current utility. Second, by running a series of artificial data experiments with a flexible specification on the functional form of the storage costs, and showing that we can recover the discount factor.

We note that our strategy to identify the consumer discount factor will work well for many, but not all product categories. Product categories where it will work well will have three key features. First, they should be product categories where a consumer does not get a large current payoff from consuming beyond weekly needs. Products such as laundry detergent or ketchup will fit this criterion well. One does not gain utility from consuming more laundry detergent than what is necessary to do laundry, or more ketchup than what is necessary to put on a hamburger. Products where temptation is a large part of purchase, such as ice cream or potato chips, may not “provide” exclusion restrictions. The reason for this is that the more of the product one has in
inventory, the more one is tempted to consume the product, and the more one gains in current utility. The second key feature is that the cost of storing a product (in terms of space used) does not in general change as inventory drops. This feature will exist in product categories where a product’s package size does not decrease with inventory - outside of rare instances where one has multiple packages and a package is used up, the space taken up by packages won’t change as the amount in a package changes. For products such as laundry detergent, where the product is a liquid stored in bottles, this assumption will hold: the cost of storing the product only depends on the number of bottles held, but not the amount of inventory within a bottle. If this were not the case, we would not have exclusion restrictions because inventory would affect storage costs continuously, which are part of a consumer’s current period payoffs. Product categories we think would work well with our identification strategy include laundry detergent, ketchup, cereal, deodorant, facial tissue, household cleaners, mustard, mayonnaise, margarine, peanut butter, or shampoo. The third key feature is that the consumption rate is sufficiently low that it takes consumers several periods to use up a package. As we will argue throughout the paper both the slope and curvature of the purchase hazard will help identify the discount factor. If consumers use up a package of the product very quickly then the purchase hazard will have little to no curvature, and it will provide less information about how forward-looking consumers are.

The final exercise we perform is to apply our approach to field data. We estimate a dynamic structural model of stockpiling behavior on IRI scanner data for laundry detergents, allowing for continuously distributed unobserved heterogeneity in most of the model parameters, particularly the discount factor. We find that consumer discount factors lie between about 0.85 and 0.96, and average about 0.91. The values of the discount factors for most consumers are significantly lower than the value of 0.95 or 0.99 that many papers assume when estimating dynamic discrete choice models of consumer behavior. This result could have strong substantive implications in the answering the questions that the literature has examined (e.g., short-term vs. long-term responses to temporary and permanent price cuts).

An outline of the paper is as follows. In Section 2, we discuss related work. Section 3 introduces a simple stylized model of stockpiling behavior, and Section 4 contains proofs of the important properties of the model. Section 5 presents conditions for identification of the discount factor when inventory is observed. Section 6 describes how the discount factor can be identified when inventory is unobserved. Section 7 describes the results of our artificial data experiments. Section 8 describes our empirical application and the estimates, and Section 9 concludes.

Interestingly, our results are consistent with the field work of Yao, Mela, Chiang, and Chen (2012), who find consumer daily discount factors are around 0.91, lower than the rational expectations benchmark.
2 Review of Literature

Proofs of identification of the discount factor often build on the conditional choice probability approach introduced in Hotz and Miller (1993). In the Hotz and Miller (1993) approach, the researcher assumes that the same state variables that are observed to the consumer 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 flexibly estimate a consumer’s choice specific value, which is the sum of the current period flow utility and the discount factor multiplied by the value function. The choice specific values are identified conditional on a normalization of the utility of one alternative (typically called the reference alternative), and given the functional form of the error distribution. With no restrictions on the functional form of the flow utility, the discount factor is not identified: in the conditional choice probability approach, one can think of each estimating equation as the probability of a consumer choosing each alternative at each value of all the state variables. A fully flexible model would allow the utility function to have a parameter that was unique for each alternative and each state. Hence, if the discount factor were fixed the number of equations and unknowns would be equal, and the model would be exactly identified. Formally, to identify the discount factor, some restriction must be put on the functional form of the utility function. Such a restriction will reduce the number of parameters in the model to be smaller than the number of equations, allowing the discount factor to be identified.

One such type of restriction that has been proposed to help identify the discount factor in this setting is called an exclusion restriction. Magnac and Thesmar (2002) first proposed the exclusion restriction approach, in the form of a functional restriction on the choice specific values. Fang and Wang (2015) build on the approach of Magnac and Thesmar (2002) and show that one can identify the discount factor in the conditional choice probability setting if one can identify two values of the state variable where, for each alternative, flow utilities are the same for both values, but the value functions differ. If this type of restriction is imposed on the model, then the discount factor can be identified. We note that technically, any restriction which reduces the number of model parameters below what is required for exact identification (in a setting where the discount factor is fixed) will enable the discount factor to be identified. For instance, one could assume that utility functions were linear or quadratic in state variables. The appeal of an exclusion restriction is that it can arise as a result of institutional features of the market under examination, rather than resulting from potentially ad-hoc functional form assumptions.

To our knowledge, the exclusion restriction approach has not been applied to the stockpiling setting. An alternative, and complementary, approach to identifying the discount in stockpiling markets has recently been developed by Akça and Otter (2015). Their approach relies on three key
assumptions: First, that brands within inventory are consumed according to a Last-In-Last-Out (LILO) rule. The idea behind this rule is that consumers keep track of when they purchase each brand of a product, and know approximately when they will consume each brand in their inventory. The most recently purchased brand will be the last product consumed in the queue. The second assumption is that the researcher observes consumer inventory. This assumption is key since the approach of Akça and Otter (2015) relies on conditional choice probabilities to prove identification. Finally, the proofs of identification rely on the error term taking on an extreme value type 1 (logit) distribution. The Akça and Otter (2015) approach is complementary to our work in the sense that it operates under a different set of assumptions. In product categories and data where inventory can easily be observed or imputed by the researcher, and the LILO assumption is likely to be true, the Akça and Otter (2015) approach can be applied. Our approach could be used in situations where the LILO assumption might be violated, or where the researcher does not observe inventory. To apply our approach however, one would need to be comfortable assuming the exclusion restriction.

A final and related work we wish to discuss is a theory testing paper by Hendel and Nevo (2006b), which proposes a series of tests for the presence of forward-looking behavior in storable goods markets. The paper develops a stockpiling model with endogenous consumption from inventory, and where consumers are able to purchase quantities in continuous amounts. In their setting, the key difference between a myopic consumer and a forward-looking consumer is that a myopic consumer will always purchase exactly the amount she will consume in the period where the purchase occurs, while a forward-looking consumer will purchase for future consumption. This type of analysis will apply well to settings where consumers have the ability to purchase the product category in small increments: for example, canned tuna or soup. A key difference between Hendel and Nevo (2006b)’s setting and ours is that in our setting myopic consumers will purchase more than they can consume in a single period, since package sizes are large relative to consumption rates. We rely on exclusion restrictions to separate out myopic consumers from forward-looking consumers, rather than relying on identification from quantity purchased. Because we rely on exclusion restrictions, we can identify the discount factor in situations where consumers are only able to purchase a single package in a purchase occasion. Another key difference is that in our setting we assume consumption rates are exogenous, in the sense that consumers use enough of a product to satisfy an exogenous consumption need (for example, one does not get extra utility from consuming more laundry detergent than is needed to do the weekly laundry). We note that the exclusion restriction may be violated in a setting where consumption is endogenous, since optimal consumption (and hence flow utility) can be a function of inventory.
3 A Stylized Stockpiling Model

In this section we describe a model that is simplified somewhat from the model we will use for our empirical application, but contains its most important features. The econometrician observes a market containing $N$ consumers making purchase decisions over $T$ periods. Consumers are forward-looking and discount the future at a discount rate $\beta_i < 1$. In this stylized model, we assume that a single product is available to consumers in some discrete package size. Each decision period $t$ is broken up into two phases: a purchase phase and a consumption phase. In the purchase phase, consumer $i$ observes the price of a package of the product, which we denote $p_{it}$, an exogenous consumption need $c_{it}$, and a choice-specific error $\varepsilon_{ijt}$. The consumer’s choice is her decision of how many packages of the product to buy, which we denote as $j \in \{0, 1, ..., J\}$. After making her purchase, the consumer receives her consumption utility.

We denote the size of a package as $b$, and for simplicity of exposition we assume that $b$ is an integer (we will relax this assumption in the empirical application). We denote the consumer’s inventory (which will also be integral) at the beginning of the period as $I_{it}$. Consumption rates $c_{it}$ will be in the set $\{0, 1, 2, ..., \bar{c}\}$. If the consumer’s inventory at the end of the purchase phase, which we denote as $I_{it} + b \cdot j$, is above the consumption need $c_{it}$ then she receives consumption utility $\gamma_i$. If she cannot cover her consumption need then she incurs a stockout cost $\nu_i$. At the end of the period the consumer incurs a storage $s(\cdot; \omega_i)$. Here we formally introduce our first assumption about $s(\cdot; \omega_i)$, which would allow the inventory variable to generate exclusion restrictions.

**First Model Assumption Related to Exclusion Restrictions, X1**

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

The number of packages held can be written as the following function of inventory $B_{i,t+1}(j, c_{it}) = \lceil (I_{it} + b \cdot j - c_{it})/b \rceil$. The assumption that $b > 1$ ensures that X1 is meaningful. $\omega_i$ is a vector of parameters determining how storage costs vary with the number of packages held. We will parameterize the storage cost function as flexibly as possible:

\[ s(B; \omega_i) = \omega_{i,B}. \quad (1) \]

---

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

5 The ceiling function $\lceil \cdot \rceil$ returns the smallest integer that is greater than or equal to its argument.
This functional form is nonparametric in the sense that there is a different parameter, \( \omega_{i,B} \) for each possible number of packages held. In practice, one may consider imposing a functional form on \( s \), such as quadratic. We will assume that the cost of storing 0 packages is 0.

The assumption that a consumer’s storage cost depends on the number of packages held of a product is going to be true for many product categories. For example, products that are sold in bottles such as laundry or dish detergent will satisfy this assumption. The cost to storing these types of products depends on the amount of space taken up by the bottle, but not the amount of liquid within the bottle. Products where a package can be compressed as the product gets used up, like potato chips, would likely not satisfy the assumption. Additionally, identification could be complicated when the consumption rate is close to the package size. In this case inventory would always be close to or equal to the number of packages held. Thus identification could be difficult in categories like canned tuna or canned soup.

Denote the volume or size of a package as \( b \), and \( I_{it} \) as the consumer’s inventory at the beginning of the period. If her inventory of the product at the end of the purchase phase, which we denote as \( I_{it} + b \cdot j \), is above the consumption need \( c_{it} \) then she receives consumption utility \( \gamma_i \). If her inventory is 0 then she incurs a stockout cost \( \nu_i \). At the end of the period the consumer incurs a storage cost, \( s(B_{i,t+1}(j, c_{it}); \omega_i) \), where \( B_{i,t+1}(j, c_{it}) = [(I_{it} + b \cdot j - c_{it})/b] \) is the number of packages held by the consumer at the end of the period and \( \omega_i \) is a vector of parameters. The assumption that a consumer will consume up to her consumption need is important, as it implies that flow utility will not depend on inventory. To emphasize this we formalize this in the following assumption:

**Second Model Assumption Related to Exclusion Restrictions, X2**

2. Consumption rates are exogenous: a consumer receives no utility from consuming more than her consumption need, and will not consume less if enough inventory is available.

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

\[
\begin{align*}
\psi_{it}(j, I_{it}, \varepsilon_{ijt}, p_{it}, c_{it}; \theta_i) &= \begin{cases} 
\gamma_i - s(B_{i,t+1}(j, c_{it}); \omega_i) - \alpha_i p_{it} j + \eta \varepsilon_{ijt} & \text{if } I_{it} + b \cdot j \geq c_{it} \\
-\nu_i - \alpha_i p_{it} j + \eta \varepsilon_{ijt} & \text{otherwise}
\end{cases}.
\end{align*}
\]

The vector \( \theta_i = (\alpha_i, \beta_i, \gamma_i, \nu_i, \omega_i) \) is a vector of the consumer utility coefficients and the discount factor. The parameter \( \alpha_i \) is the price coefficient. The parameter \( \eta \) is an error term weight that will be relevant for some of the theoretical proofs. In estimation we will normalize this parameter to 1.

We assume that consumers believe that the product’s price follows a stochastic Markov process with a transition density \( F(p_{i,t+1}|p_{it}) \). The consumption shocks \( c_{it} \) are i.i.d over time for each
consumer and are drawn from a discrete distribution where the probability of receiving consumption shock level \( l \) as \( \pi^l_c \). The consumer’s Bellman equation is as follows:

\[
V_{it}(I_{it}, p_{it}) = \sum_{l=0}^{\infty} \mathbb{E}_{\varepsilon_{it}} \max_{j=0, \ldots, J} \{ u_{it}(j, I_{it}, \varepsilon_{ijt}, p_{it}, l; \theta_i) + \beta_i \mathbb{E}_{p_{it+1}} V(I_{it+1}, p_{it+1}) \} \pi^l_c. \quad (3)
\]

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

\[
I_{i,t+1} = \max\{I_{i,t} + b \cdot j - c_{it}, 0\}.
\]

We also put an upper bound on the number of packages a consumer can carry, which we denote \( M \). We assume that if a consumer makes a purchase when her inventory is above \( M - b - c_{it} \), then her inventory is set to the upper bound \( Mb \). Intuitively this is consistent with a situation where a consumer’s storage space is used up, but if she purchases another bottle she takes the one that is already open and gives it away or otherwise disposes of it.

4 Model Properties

In this section we will derive some useful properties of the model above, which will help us to understand intuitively what type of variation in the data will help identify the discount factor. The basic idea behind the identification of \( \beta \) is that there are exclusion restrictions in our model: Consumer inventory enters the expected future payoffs in a continuous way, and it almost never directly affects consumer’s current payoffs.\(^6\) Intuitively, unless a consumer is very close to using up a bottle/package of laundry detergent, washing an extra load of laundry will lower the inventory level, but it does not change the storage cost. Therefore, if a consumer cares about her future payoffs, her incentive to avoid the stock out cost \( \nu \) also gets stronger and stronger if she is more forward-looking (i.e., \( \beta > 0 \)). In other words, the functional relationship between consumer’s purchase incidence and inventory should depend on the value of her discount factor. Therefore, a consumer’s discount factor should be identified if we observe consumer’s choice at different inventory levels.

\(^6\)As we describe in more detail below, the two exceptions are at the exact point a consumer stocks out, and at the point a consumer uses up a package. Since these are rare events we argue they should not affect the identification of the discount factor.
Formally, we can express the above intuition by deriving two key properties of the value function. First, the expected future value of a purchase should increase as inventory drops, at least for sufficiently low values of inventory. Second, as consumers get more forward-looking, the expected future value of a purchase should rise. We will prove these statements are true in simple setting with no price variation, and where the error term is has zero variance ($\eta = 0$). We will then demonstrate that the value function is continuous in $\eta$ under some regularity conditions on the error term, and if payoffs can be bounded, which will demonstrate that for small values of $\eta$ the same properties of the value function will hold.

We will make some simplifications to the model which we will relax in later sections. In the analysis below we will normalize the price coefficient, $\alpha$, to 1. For the theory this is innocuous: we could include this parameter, and scale prices by a factor of $\alpha$, and our results would be unchanged. We will also assume that all the model parameters are homogeneous across the population, and thus will drop the $i$ subscript on everything except the state variables and the error term.

The assumptions we will make to prove the properties of the value function we want are listed below.

**Assumptions A1-A8**

1. Consumption shocks are in the set \{1, 2\}. The probability that $c_t = 1$ is $\pi_c$.

2. The maximum number of packages that can be purchased in a period is $J = 1$.

3. Prices are fixed over time at a level $p > 0$.

4. The package size $b$ is greater than or equal to 2.

5. The weight on the error term, $\eta$, is small.

6. The stockout cost $\nu$ is strictly positive.

7. $\omega_1$ is small, and $p < \nu - \omega_1$.

8. The storage cost function is weakly increasing and weakly convex.

**Proposition 1** If assumptions A1-A8 hold, and $\eta = 0$, then the expected future value of making a purchase, $\beta(V(I + b - c) - V(I - c))$, is decreasing in $I$ for $I \geq c$.\(^7\) It is strictly decreasing if $\beta > 0$ and 0 if $\beta = 0$.

\(^7\)If the convexity of the storage cost function is violated the proposition still holds, but only for $I \in [0, b - 1]$.
Proof. The steps we follow are first to propose an optimal policy to the consumer’s problem, to derive the properties of the value function under that policy, and then to show the policy solves the Bellman equation. The proof of the proposition will result.

First, note that the optimal policy for all $\beta \geq 0$ is for a consumer to purchase only when she runs out. This means a purchase only occurs if the consumer starts the period with 0 inventory or 1 inventory and gets a consumption shock of 2. We note that since there is no error term in this model a myopic consumer and forward-looking consumer will behave in the same way. A myopic consumer will always find it optimal to purchase rather than stocking out due to the assumption that $p + \omega_1 < \nu$. In the presence of error shocks or varying prices forward-looking consumers will not follow the myopic strategy.

In Appendix 11.1 we prove the following Lemma:

**Lemma 1** The value function $V(I)$ is increasing in inventory for $I < b + 3$ and sufficiently small $\omega_1$.

This lemma says that a consumer’s future value rises is inventory rises. This will hold as long as the consumer’s future value does not involve an increase in storage costs beyond 1 package.

We can use Lemma 1 to prove the Lemma 2, that the proposed policy is optimal.

**Lemma 2** For all $\beta > 0$, it is optimal to purchase only when $I - c < 0$.

The proof of Lemma 2 proceeds by using induction, and is shown in Appendix 11.2.

With Lemma 2 in hand we can prove the proposition. The inequality we wish to prove is

$$V(I + 1 + b - c) - V(I + 1 - c) < V(I + b - c) - V(I - c).$$

Consider the case where $I - c = 0$. Using Lemma 2 we can show that

$$V(b) - V(0) = p$$

$$V(b + 1) - V(1) = (1 - \pi_c(1 - \beta))p - \pi_c\omega_1$$

It is clear from the equations above that $V(b + 1) - V(1) < V(b) - V(0)$ since $p > 0$, $0 \leq \beta < 1$, and $0 \leq \pi_c \leq 1$. We also need to show that $V(b + 2) - V(2) < V(b + 1) - V(1)$. We can use the fact that $V(2) = \beta \pi_c V(1) + \beta(1 - \pi_c) V(0) - \pi_c \omega_1$ to show that $V(b + 2) - V(2) = p(\beta \pi_c (1 - \pi_c (1 - \beta) + \beta (1 - \pi_c))) - (1 + \beta \pi_c^2 - \pi_c) \omega_1$. It is sufficient to show the inequality

$$V(b + 2) - V(2) < V(b + 1) - V(1).$$
\[ \beta \pi_c (1 - \pi_c (1 - \beta)) + \beta (1 - \pi_c) < 1 - \pi_c (1 - \beta) \]

\[ \iff \beta - \pi_c (1 - \beta)(1 - \pi_c \beta) < 1 \]

The second inequality will always be true since \( \beta < 1 \).

Now we use the induction step. Suppose that \( V(n + b) - V(n) < V(n + b) - V(n) \) for all \( n = 0, ..., I + 1 \) (we can subsume the \( c \) into \( I \) since it occurs in all the arguments of \( V \)). There are a few cases we want to consider. Suppose that there is no storage cost change between \( I + 2, I \), and \( I - 1 \). In that case the inequality we want to prove is

\[ V(I + 2 + b) - V(I + 2) < V(I + 1 + b) - V(I + 1) \]

\[ \iff \pi_c \beta (V(I + 1 + b) - V(I + 1)) + (1 - \pi_c) \beta (V(I + b) - V(I)) \]

\[ < \pi_c \beta (V(I + b) - V(I)) + (1 - \pi_c) \beta (V(I + b - 1) - V(I - 1)) \]

The second inequality in the above equation will be true due to our induction assumption. This proves the proposition up to the level \( I + 1 = b \).

Now suppose we observe a storage cost change between \( I + 1 \) and \( I \), or \( I \) and \( I - 1 \). Then, in addition to the above inequality, it will need to be the case that

\[ \omega_B - \omega_{B-1} \geq \omega_{B-1} - \omega_{B-2}, \]

which is just convexity of storage costs. We note that this is an assumption that is commonly made in prior research on stockpiling. ■

The second proposition we want to prove is that as \( \beta \) rises, the increase in the expected future payoff from making a purchase goes up as inventory goes down for inventory that is below some bound \( \bar{I} \). We show this in the following proposition:

**Proposition 2** If assumptions A1-A8 hold, \( \eta = 0 \), then the expected future value of purchase from an increase in inventory, \( \beta [V(I + b) - V(I)] \), is strictly increasing in \( \beta \) for inventory \( a \) for a range \( I \in [0, \bar{I}] \), where \( \bar{I} \geq 0 \).

**Proof.** First we note that this payoff will be 0 if \( \beta = 0 \). We first want to show there is some area of the state space where \( V(b + I) - V(I) > 0 \). In this area, it is sufficient to show that \( V(b + I) - V(I) \)
is weakly increasing in $\beta$ for the increasingness result to go through. We will use induction to do this.

The base cases are $V(N) − V(0)$ and $V(N + 1) − V(0)$. We know that $V(N) − V(0) = p > 0$, so it must be that

$$\frac{∂V(N) − V(0)}{∂β} = 0.$$ 

Note that this shows that $\bar{T} ≥ 0$. We also know that $V(N) − V(0) = p > 0$, so it must be that

$$\frac{∂V(N) − V(0)}{∂β} = 0.$$ 

The derivative of this with respect to $β$ is $πcP ≥ 0$. As long as $(1 − πc(1 − β)p − (1 − πc)ω1 ≥ 0$ then the derivative of $β(V(b + 1) − V(1))$ will be increasing.

Now do the induction step. Suppose that $V(N + n) − V(n)$ is weakly increasing in $β$ for all $n ≤ I$, and additionally that $V(N + n) − V(n) > 0$ for all points prior to $I$. Then let’s look at $I + 1$. We can write

$$V(N + I + 1) − V(I + 1) = βπc(V_{N+I} − V_I) + β(1 − πc)(V_{N+I−1} − V_{I−1}) + Δω,$$

where $Δω$ is a difference in storage costs (which is not a function of $β$). Both the terms $V_{N+I} − V_I$ and $V_{N+I−1} − V_{I−1}$ are increasing in $β$ by the induction assumption. This implies that $V(N + I + 1) − V(I + 1)$ is increasing in $β$. As long as $V(N + I + 1) − V(I + 1) > 0$, $β(V(N + I + 1) − V(I + 1))$ will also be increasing in $β$.

Note that the interval over which the expected future payoff is increasing, $[0, \bar{T}]$, will rise storage costs decrease. If there are no storage costs then it can be shown that $\bar{T} = \infty$, because in that case the value function is strictly increasing in inventory, meaning the expected future payoff from purchase is always positive.

The next proposition shows continuity. Here we make some regularity assumptions on the error term, and put boundedness and sign restrictions on the payoffs:

**Assumptions E1-E2**

1. Continuity and support: The CDF of the difference in $ε_1 − ε_0$, $F$, is continuous, strictly increasing, and has support $(-∞, ∞)$.

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

   $$I − c ≥ 0 : \; (−P − (ω_{B+1} − ω_{B}1{I > c}) + β(V(I + b − c) − V(I − c)) < 0$$

   $$I − c < 0 : \; (−P − (ω_{B+1} − ω_{B}1{I > c}) + β(V(I + b − c) − V(0)) > 0$$
The above assumptions should imply the following proposition:

**Proposition 3** If assumptions A1-A4 and E1-E2 hold then the expected future value of purchase from an increase in inventory, \( \beta [V(I + b) - V(I)] \), is continuous in \( \eta \).

**Proof.** To see this note that for \( I \geq 2 \) the probability of a purchase can be written as

\[
P(I, c) = F((-p - (\omega_{B+1} - \omega_B I > c}} + \beta(V(I + b - c) - V(I - c))/\eta)
\]

The value function for \( I \geq 2 \) can be written as

\[
V(I) = \pi_c (P(I, 1)(-p - \omega_{B+1} + \beta V(I + b - 1)) + (1 - P(I, 1))(-\omega_B I > 1} + \beta V(I - 1))
\]

\[
+ (1 - \pi_c)(P(I, 2)(-p - \omega_{B+1} + \beta V(I + b - 2)) + (1 - P(I, 2))(-\omega_B I > 2} + \beta V(I - 2)/\eta)
\]

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

\[
\lim_{\eta \to 0} P(I, c) = 0,
\]

and otherwise

\[
\lim_{\eta \to 0} P(I, c) = 1.
\]

It is clear that the limit as \( \eta \) approaches zero of equation (4) will equal the value function that is obtained when \( \eta = 0 \), which we derive in the proofs of Lemma 1 and Lemma 2. Similar findings will be obtained for the value function when \( I = 0 \) or \( I = 1 \).

Propositions 1 and 2 are important because they imply that a forward-looking consumer will, at least for some of the state space, be more likely to purchase than a myopic consumer. More importantly, the likelihood a forward-looking consumer purchases will increase as inventory drops, since the future value of purchase is decreasing in inventory. Although we have found a proof of Propositions 1 and 2 to be feasible in the case with no error term, Proposition 3 shows that there is a class of models including an error term where the expected future value of purchase will be decreasing in inventory, and increasing in the discount factor. To provide some more intuition in Figure 1 we plot the expected future value of a purchase for different values of the discount factor, for a low and high value of the stockout cost \( \nu \). For a forward-looking consumer, the expected future value of a purchase rises as inventory drops because purchasing delays the likelihood of a future stockout. Additionally, when inventory is sufficiently low, increasing the discount factor increases
the expected future payoff from purchase. When storage costs are positive and the stockout cost is low, at sufficiently high levels of inventory the expected future value of purchase can be decreasing in the discount factor, since adding inventory will increase storage costs in the future, which will counterbalance the gain from delaying the stockout cost.

Figure 1: Expected future payoff from purchase, $\beta(V(I + b) - V(I))$, as a function of $I$ and $\beta$. Parameter values $\pi_c = 0.5, \omega_1 = 0, \omega_2 = 0.05, \omega_3 = 0.15, \eta = 1, M = 3, p = 2$, and logit error term.

5 Identification with Observed Inventory

In this section we will discuss how the discount factor can be identified in the situation in which inventory is observed. Although in most empirical applications inventory will be unobserved, we feel that understanding the features of the model that drive identification in this setting will help the reader to understand what drives identification in the setting where inventory is not observed. For this section and for Section 6 prior to Section 6.4 we will maintain the assumption that prices do not vary. As we will show in Section 6.4, price variation will aid identification in the sense that it accentuates intertemporal tradeoffs. Another simplification we have made is that consumers cannot purchase more than a single package. As with price variation, being able to purchase more packages will help us to identify the discount factor, which we will address in Section 6.4. The fact that we can identify the discount factor in a setting with no price variation and where consumers only purchase a single package is encouraging, as it suggests that one can identify forward-looking behavior in situations where the data might initially look too limiting to facilitate identification.

One other assumption we will maintain is that inventory is integral. Intuitively, identification in a setting where inventory is continuous will not be that different from the discrete setting. As
long as the exclusion restrictions hold, there will be some areas of the state space where current utility does not change but the expected future payoff does.

Below we provide a set of assumptions which are necessary for identification in the formulation of the stockpiling model we have been examining. For identification, we can weaken some of the assumptions we needed to prove the properties of the model. At the end of the section we will discuss ways in which the assumptions can be weakened further.

**Assumptions A1’-A7’**

1. Consumption shocks are in the set \( C = \{0, 1, \ldots, \bar{c}\} \), the distribution of shocks is i.i.d. and discrete.

2. The maximum number of packages that can be purchased in a period is \( J = 1 \).

3. Prices follow a discrete Markov process, taking on \( L \) values \( p_0, \ldots, p_L \).

4. The package size \( b \) is greater than or equal to \( \bar{c} \).

5. The weight on the error term is strictly positive: \( \eta > 0 \).

6. The stockout cost is strictly positive: \( \nu > 0 \).

7. Inventory is observed to the researcher. The researcher observes consumer choices for at least one price level \( p_l \), and for at least the following set of consumption and inventory shock pairs:

\[
(I_1, c_1) \in C_1 = \{(0, c) : c \in C\}
\]

\[
(I_2, c_2) = C_2 = \{(1, 1)\}
\]

\[
(I_3, c_3) \in C_3 = \{(I, c) : I - c \geq 1 \text{ and } I - c < 2b + 1, c \in C\}
\]

\[
(I_{3+i}, c_{3+i}) \in C_{3+i} = \{(I, c) : I - c > (i + 1)b \text{ and } I - c < (i + 2)b + 1, c \in C\}, i = 1, \ldots, M - 2
\]

\[
(I_{2+M}, c_{2+M}) \in \cup_{i=3}^{M-2} C_i, I_{2+M} - c_{2+M} \neq I_3 - c_3, \ldots, I_{M-1} - c_{M-1}
\]

We also add the following rank condition:

**Rank Condition R1**

1. Define the vector
The above system will be invertible when rank condition R1 holds. □

The arguments and conditions we use in this section parallel those proposed in Fang and Wang (2015). If inventory is observed then the researcher can estimate choice probabilities at each level.
of inventory, and invert the choice probabilities to recover the choice-specific values, \( \hat{v}_i \), which are used to solve for the model parameters. Intuitively, we can think of observations with 0 inventory as being the ones that identify \( \nu \), observations with 1 inventory as identifying \( \omega_1 \) (since at these observations a consumer will not run out, but will incur a storage cost if she purchases), and observations with inventory above 1 where storage costs do not change as being the ones that identify \( \beta \). We assume that there is at least one pair of inventory levels where this occurs: \( I_{2+M} \) and one of the other levels. Suppose we denote the number of bottles where inventory level is \( I_{2+M} - c_{2+M} \) as \( B - 1 \), and denote as \( i \) the other inventory level where the number of bottles held at the end of the period is \( B - 1 \) if no purchase is made. Then for these two states, notice that the difference between choice specific values for inventory \( I_{2+M} \) and \( I_i \) are simply

\[
\beta \left( V(\min\{Mb, b + I_{2+M} - c_{2+M}\}) - V(I_{2+M} - c_{2+M}) - (V(b + I_i - c_i) - V(I_i - c_i)) \right) = \hat{v}_{2+M} - \hat{v}_i.
\]

By assumption, the storage costs at these state points are the same, but the states are different so any difference in choice specific values must be due to difference in the value functions. This is exactly the exclusion restriction at work. The baseline purchase probabilities at the other inventory levels \( I_i \) indexed by \( i = 3, ..., 1 + M \) will identify the rest of the storage cost parameters.

To illustrate this intuition we plot the purchase probabilities as a function of time for different discount factors in Figure 2, in the following situation: We suppose that in period 0 the consumer starts with 2 full packages of the product, and her consumption shocks are 1 each period. In each period, the consumer’s inventory drops by 1 unit, until period 17, where she runs out. Consider first the black line, which shows the probability of purchase for a completely myopic consumer. This consumer’s purchase probability is flat except at 3 periods: period 8, where a package is used up and the storage cost drops, period 16 where the storage cost is 0 at the end of the period, and period 17 where the consumer runs out. For most of the state space the purchase probability is flat, since the flow utility is not changing as a result of the exclusion restrictions. Now compare the purchase probability of a forward-looking consumer. For a forward-looking consumer, the purchase probability rises over time in the areas where it is flat for the myopic consumer. This occurs because the future value of purchase decreases as inventory rises, as shown in Proposition 1. Additionally for the range of time where the consumer has only a single package left this increase will be larger for larger values of \( \beta \). This is a result of Proposition 2, which implies the value of purchase should rise for sufficiently low levels of inventory.

Note that the level of the purchase probability depends on both the storage cost and the discount factor. If inventory is at a level where storage costs are higher, the purchase probability can be lower than the myopic case because future storage costs are included in the choice specific value. The important point is that the slope and shape of the purchase probability in areas where the
exclusion restrictions hold will identify the discount factor. Proposition 2 also helps us understand how the discount factor can be uniquely identified, in the sense that for periods before someone runs out, the purchase probability is strictly increasing in the discount factor (this is a result of the fact that the expected future payoff increases in the discount factor). Given a particular value of \( \nu \), we can map the purchase probability prior to running out into a unique value of the discount factor. Note that we do have a failure of identification if \( \nu = 0 \) and all the storage costs are zero. In this case a myopic consumer payoffs will be the same as a static consumer’s payoffs. Intuitively, there is no value to stockpiling, and no discounting of future storage costs. Assumption A6’ guarantees this does not happen.

A final note is that Assumptions A1’-A7’ are tailored to produce identification in the particular formulation of the model we have been using. It is easy to see that some of these assumptions could be relaxed. For instance, A2’ could be relaxed to allow for choices of multiple sizes. One would also need to modify the regularity conditions on the error to allow invertibility. Such conditions are outlined in Hotz and Miller (1993). Furthermore, one could allow for continuous consumption rates, continuous inventory, and continuous prices. The important point is that one observes inventory held near 0, at all possible levels of storage costs, and that for one level of storage costs 2 different inventory points are observed, in order to allow identification of the discount factor.

6 Identification with Unobserved Inventory

In this section we consider identification with unobserved inventory. Scanner datasets that are often used by researchers to estimate stockpiling models do not track consumer inventory. We will argue that identification of the discount factor will arise from the shape and slope of the purchase hazard, which has similar properties to the purchase probability shown in Figure 2. The purchase hazard is the average probability of purchase in period \( t + \tau \) given a purchase was observed in period \( \tau \). The complication that we face is that we only know that inventory rose by an amount \( b \) in period \( t \), but we do not know inventory at the time of purchase and must integrate it out. Additionally, to compute the probability of purchase conditional on the last purchase occurring \( t \) periods in the past we must integrate out of the intervening sequence of consumption shocks. In Section 6.1 we will provide formal conditions for identification based on the purchase hazard and observed purchase frequency. Identification will rely on a rank condition which may be difficult to verify, since the purchase hazard is a more complicated function of the model’s parameters than the choice probability conditional on inventory. We will argue that assumptions X1 and X2, plus a restriction on the consumption rate distribution, can help ensure the rank condition will hold. We
will then argue using simulations and numerical solution of the model that the discount factor is likely to be identified when these assumptions are true.

In the numerical solutions below we will assume that the package size, \( b \), is 8 units, and consumption shocks are in the set \{1, 2\}. We will allow consumers to store up to \( M = 3 \) packages and will assume the error term is standard logit. When we compute the purchase we will need to simulate out the steady state distribution of inventory in the population. To do this we will simulate purchases for 500 individuals for 600 periods. We will assume that in period 0 all individuals have 0 inventory. We find that aggregate inventory appears to reach the steady state at around 50 periods, so we will use periods 400 to 600 to compute steady state inventory. We will also assume (prior to Section 6.4) that prices are fixed over time at a level of 2.

### 6.1 Identification Assumptions

In this section we formalize the assumptions necessary for identification of \( \theta = (\nu, \beta, \omega_1, \ldots, \omega_M, \pi_c) \) in the presence of unobserved inventory. Define the steady state distribution of inventory as \( \pi^I(\theta) \).
Let the indicator \( d_t = 1 \) if a purchase occurs, and 0 otherwise. Then, define the purchase hazard, which is the probability of a purchase occurring in period \( t + \tau \) given no purchases in the intervening periods, conditional on observed inventory level \( I \) in period \( t \) and an observed price history \( p_t, ..., p_{t+\tau} \), as

\[
\phi_{\tau}(\theta; I, p_t, ..., p_{t+\tau}) = \text{Prob}(d_{t+\tau} = 1 | d_t = 1, d_{t+1} = 0, ..., d_{t+\tau-1} = 0; I_t = I, \theta, p_t, ..., p_{t+\tau}).
\] (6)

Note that to construct the probability in equation (6), we must integrate out over the sequence of consumption shocks that occur in the intervening \( \tau \) periods between \( t \) and \( t + \tau \). Since inventory is unobserved, the purchase hazard cannot be computed empirically without a fully specified econometric model. However, the researcher could flexibly estimate the aggregate purchase hazard, which is the average probability of a purchase in period \( t + \tau \) given a purchase occurs in period \( t \). We define the aggregate purchase hazard as

\[
\Phi_{\tau}(\theta; p_t, ..., p_{t+\tau}) = \sum_{l=0}^{7} \pi^l(\theta)\phi_{\tau}(\theta; I, p_t, ..., p_{t+\tau}).
\] (7)

We will argue that the aggregate purchase hazard can be used to identify the parameters \( \nu, \beta, \omega_1, ..., \omega_M \). The consumption rate \( \pi_c \) can be identified from the total quantity purchased, divided by the total number of weeks where consumers are observed. This moment will theoretically be

\[
M(\theta) = b \sum_{l=1}^{L} \text{Prob}(p = p_l) \sum_{l=0}^{7} \text{Prob}(d_t = 1; \theta, I, p).
\] (8)

If a researcher observes these moments, then we can define a rank condition as follows:

**Rank Condition R2**

1. Assume that the purchase hazard can be computed for a single price history of length \( Mb \), and denote the initial period as 0. Define the vector

\[
\mathbf{v} = \begin{bmatrix}
\Phi_1(\theta, p_0, p_1) \\
\vdots \\
\Phi_{Mb}(\theta, p_0, ..., p_{Mb}) \\
M(\theta)
\end{bmatrix}
\]

The Jacobian of \( \mathbf{v} \) with respect to \( (\nu, \beta, \omega_1, ..., \omega_M, \pi_c) \) is full rank.

The parameter vector \( \theta \) will be identified if rank condition R2 holds.
Proposition 5 If Assumptions A1’-A7’, Error Condition E1, and Rank Condition R2 hold then the model is identified.

Proof. The proof is a straightforward application of the Implicit Function Theorem. ■

We will argue that assumptions, X1 and X2, coupled with an additional restriction that guarantees that it takes consumers several periods to run out of a package, can help guarantee that rank condition R2 holds. Formally, the two restrictions are

Identification Restrictions I1-I2

1. Assumptions $X_1, X_2$ hold.

2. Consumption rate restrictions:

   (a) The probability that $c = 0$ is strictly less than 1.

   (b) Suppose that a purchase of a single package is observed in period $t$. Define $t + \tau$ as the first period after $t$ where the average predicted amount of inventory held reaches zero. It is the case that $\tau > M$.

Intuitively, X1 and X2 reduce the number of storage cost parameters to at most $M$, while the number of moments used to identify the model is $Mb + 1$. Since $b > 1$ it will be the case that $Mb + 1 \geq M + 2$, which is the number of parameters. Additionally, if $b > 1$ then it is possible for the exclusion restriction to hold, in the sense that we can observe consumers at two different inventory levels, having the same storage costs. Restriction I1 would be sufficient for identification if inventory were observed. In the case of unobserved inventory, we also will want to impose Restriction I2. The intuition for why we need this restriction follows from the fact that we are using the purchase hazard as a proxy for the probability of purchase given a particular inventory level. Restriction I2 will guarantee that the purchase hazard does not flatten out and become degenerate too quickly. For example, suppose that $b = 2$, and $c = 2$ with probability 1. The purchase hazard will be degenerate since people will run out in the first period after making a purchase, and will provide no identifying information (the purchase hazard will be flat since all consumers will have 0 inventory after 1 period. Conditional on having 0 inventory, the probability of purchase does not change). Restriction I2 guarantees that the purchase hazard will be increasing for at least $\tau$ periods. Intuitively we want to think of the periods 1 to $\tau + 1$ as helping to identify the discount factor and the storage cost parameters. We should observe at least $M + 1$ such periods to guarantee identification. At the end of this paragraph, we provide an assumption on the consumption rate distribution that will guarantee that I2 is satisfied.

Example of a Restriction Satisfying I2
Assume that the upper bound on the consumption rate distribution, $\bar{c}$, is such that $b/\bar{c} > M$.

Note that conditions I1-I2 will be necessary, but not sufficient, for identification to hold. It may be possible for the rank condition to fail, even if we have a sufficient number of moments where consumers have positive inventory. A failure of the rank condition if I1-I2 are maintained is likely to occur only in particular, pathological parameterizations (such as where $\nu = \omega_1 = ..., = \omega_M = 0$).

We provide some evidence for this in the following sections through numerical solution of the model.

Before we conclude this section, we note that we maintained Assumptions A1’-A7’, which as we discussed in Section 5 could be relaxed. In particular, the assumption that $J = 1$ is more restrictive than necessary. If $J > 2$, one could compute several analogs to the purchase hazard in equation (7), where one conditioned on purchases of 1 package, 2 packages, etc, in period $t$, and where in period $t + \tau$ the relevant probability was the probability of purchasing 1 package, 2 packages, etc. Allowing for multiple packages would aid identification in that for each price history one could theoretically compute $J^2$ purchase hazards, drastically increasing the number of moments available for identification. The assumptions related to the discreteness of inventory or prices could also be easily relaxed, as discussed in Section 5.

6.2 No Storage Costs

We begin by considering the case where $\omega_i = 0$ for $i = 1, ..., M$. In Figure 3 we plot the aggregate probability of purchase in period $t + \tau$ given a purchase in period $t$ for different values of the discount factor. There are two principal features of the purchase hazard that help us to tell apart a myopic from a forward-looking consumer. The first is the slope of the purchase hazard in the periods immediately after a purchase occurs. For our particular parameterization, a purchase increases an individual’s inventory by 8 units. Since consumption shocks are at most 2 units, it will take someone at least 4 periods to run out and incur a stockout cost. To see the implications of this, consider the purchase hazard for a myopic individual, shown by the black line in Figure 3. For the first 3 periods after a purchase, the purchase hazard is flat, since a myopic consumer’s flow utility is fixed over this interval. In contrast, for a forward-looking consumer the purchase hazard has a positive slope over the first 3 periods. This occurs because the expected future value of purchase rises as inventory drops, as we showed in Proposition 2. The second, and more subtle difference between the purchase hazards, is that the purchase hazard becomes smoother as $\beta$ rises (note that this feature of the purchase hazard also arises in the purchase probabilities with observed inventory in Figure 2). The intuition here is that a myopic consumer is not willing to trade off future utility for current utility, so her purchase hazard will start to rise sharply at $\tau = 4$, when people in
the population start to run out. In contrast, a forward-looking consumer will be more willing to purchase early, and so the purchase hazard will be smoother for such a consumer.

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

The identification problem with unobserved inventory becomes more complicated than with observed inventory since we have to consider separate identification of the discount factor \( \beta \), the stockout cost \( \nu \), and the probability of a low consumption shock \( \pi_c \). A feature of the model that aids identification is that \( \nu \) and \( \pi_c \) have very different effects on the purchase hazard than \( \beta \). Figure 4 shows how \( \nu \) affects the purchase hazard, for low and high values of the discount factor. Most of the impact of a change in \( \nu \) on the purchase hazard occurs during later rather than earlier periods. This is sensible since \( \nu \) should have more impact on purchase decisions when consumers begin to run out. Importantly, the shape of the purchase hazard is preserved as \( \nu \) changes - for the low value of \( \beta \), the purchase hazard displays a lot of curvature around period 4 for different values of \( \nu \). Similarly, the purchase hazard is very smooth for high values of \( \beta \) for different values of \( \nu \). The impact of \( \pi_c \) on the purchase hazard for different values of \( \beta \) is shown in Figure 5. For low values of \( \beta \), the impact of changing the probability of a low shock is similar to that of \( \nu \). For high values of \( \beta \), changing \( \pi_c \) shifts the purchase hazard up and down. Our analysis does suggest that \( \nu \) and
πc could be difficult to separately identify if β is low. For this reason, we introduce additional moments to argue that we can identify the consumption rate distribution.

One way to properly identify the consumption rate distribution is to match aggregate statistics related to consumer purchases, such as the total quantity purchased by consumers divided by the total number of weeks. This measure has been used as to estimate the consumption rate in some earlier work on storable goods (need a citation here). To demonstrate that the total quantity purchased divided by total weeks can identify the consumption rate, we simulate consumer purchases for different consumption rate distributions, compute the aforementioned moment, and show the results in the second column of Table 1. The first row of the table shows that if the average consumption rate is 1.75, then the total quantity purchased per period is 1.5 units. The total quantity purchased per period is a bit below the actual consumption rate since individuals occasionally stock out. However, as the average consumption rate decreases, the total quantity purchased per period also drops. The last 3 columns of the table show how one could identify the parameters of a 3 point distribution of consumption rates. To aid the identification argument we add an additional moment, which is the average within person standard deviation of the interpurchase time (the number of periods between purchases). In rows 2 and 3, we show how the standard deviation of interpurchase time changes as the variance of the consumption rate distribution rises. As the variance goes up, the average amount purchased drops slightly, but the standard deviation in interpurchase time rises. This finding is intuitive - more variance in consumption rates should lead to higher variance in the amount of time it takes for consumers to use up their inventory, which will in turn raise the variance in the time to next purchase.

Our arguments above suggest that if the researcher can compute the purchase hazard and the average quantity purchased (and potentially the variance of interpurchase times), if these moments were generated by forward-looking consumer behavior it would not be possible to fit them under the assumption that consumers were relatively myopic, even with the right values of ν and πc. To demonstrate this we compute the purchase hazard and average quantity purchased for an individual with β = 0.99, and try to fit these moments as closely as possible by varying ν and πc while holding β fixed at 0.5. In Figure 6 we show the fitted purchase hazard (dotted line) against the actual purchase hazard for two values of ν. In each case the fit to the actual purchase hazard is poor. In particular, it is not possible for the curvature of the purchase hazard of a relatively myopic person to fit that of a forward-looking person, especially in periods around when consumers start to run out.
Table 1: Impact of Consumption Rate Distribution on Average Quantity Purchased and Standard Deviation of Interpurchase Time

<table>
<thead>
<tr>
<th>Consumption Rate Distribution</th>
<th>c ∈ (1,2)</th>
<th></th>
<th>c ∈ (0.1,2)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total Quantity/ Total Weeks</td>
<td>S.D. Interpurchase Time</td>
<td>Total Quantity/ Total Weeks</td>
<td>S.D. Interpurchase Time</td>
</tr>
<tr>
<td>(0.25,0.75)</td>
<td>1.506</td>
<td>0.743</td>
<td>(0.1,0.1,0.8)</td>
<td>1.451</td>
</tr>
<tr>
<td>(0.5,0.5)</td>
<td>1.404</td>
<td>0.823</td>
<td>(0.1,0.8,0.1)</td>
<td>1.127</td>
</tr>
<tr>
<td>(0.75,0.25)</td>
<td>1.271</td>
<td>0.965</td>
<td>(0.333,0.333,0.333)</td>
<td>1.115</td>
</tr>
<tr>
<td></td>
<td>(0.8,0.1,0.1)</td>
<td>0.948</td>
<td></td>
<td>1.674</td>
</tr>
</tbody>
</table>

Notes: This table shows how the consumption rate distribution affects two simulated moments, the first being the total quantity (purchases multiplied by package size) divided by the total number of weeks (columns 2 and 5), and the second being the standard deviation of interpurchase time (columns 3 and 6). The first 3 columns apply to a 2 point distribution of consumption rates (1 and 2), while the last 3 apply to a 3 point distribution (0, 1 and 2). The moments are computed for 500 consumers and 600 periods, where the first 200 periods are removed to reduce dependence on initial inventory. In the simulation, we fix \( \nu = 0.25, \beta = 0.95, \eta = 1, \omega_1 = \omega_2 = \omega_3 = 0, \) and the error is assumed to be logit.

Figure 4: Probability of purchase in period \( t + \tau \) given purchase of 1 package in period \( t, \) for different values of the stockout cost. Parameter values \( \pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \omega_3 = 0, \eta = 1, M = 3, p = 2, \) and logit error term.
Figure 5: Probability of purchase in period $t + \tau$ given purchase of 1 package in period $t$, for different values of $\pi_c$. Parameter values $\nu = 0.25$, $\pi_c = 0.5$, $\omega_1 = 0$, $\omega_2 = 0$, $\omega_3 = 0$, $\eta = 1$, $M = 3$, $p = 2$, and logit error term.

Figure 6: Fitted and actual purchase hazards for different true values of $\nu$. In the actual purchase hazard we set $\pi_c = 0.5$. Other parameter values $\omega_1 = 0$, $\omega_2 = 0$, $\omega_3 = 0$, $\eta = 1$, $M = 3$, $p = 2$, and logit error term.
6.3 Nonzero Storage Costs

We now analyze the case when storage costs are nonzero. With unobserved inventory, identification is complicated by the fact that when consumer are myopic, the purchase hazard may always be positively sloped. This can occur because there will be some individuals who, when they make a purchase, will have a small amount of a package left. These individuals will use up the package within a few periods after the purchase, and will observe a decrease in their storage costs. That decrease in storage costs will lead to an increase in the probability of a purchase. Because it is likely that some consumers will use up a package in every period after a purchase, in every period there will be some increase in the purchase hazard for myopic consumers. In Figure 7 we compute the purchase hazards for different discount factors, with $\omega_3 = 0.5$. The purchase hazard for a myopic consumer is positively sloped in the first 3 periods after a purchase - recall in Figure 3 a myopic consumer’s purchase hazard was flat in this region. There are two features of the purchase hazard that still hold when storage costs are positive. First, for a few periods immediately after purchase the slope of the purchase hazard is still increasing in $\beta$. Additionally, the purchase hazard becomes smoother as $\beta$ rises, especially around period 4 when consumers begin to run out.

![Figure 7: Probability of purchase in period $t+\tau$ given purchase of 1 package in period $t$. Parameter values $\nu = 0.25, \pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \omega_3 = 0.5, \eta = 1, M = 3, p = 2$, and logit error term.](image-url)

29
Figure 8 shows the impact of changing storage costs on the purchase hazard. Storage costs have little effect on the purchase hazard in later periods, as the incentive to purchase and avoid the stockout cost outweighs the incentive to avoid incurring higher storage costs. The main effect occurs early on, where consumers who already have some inventory avoid purchasing more to avoid paying a higher storage cost. To demonstrate that it is not possible to fit the observed moments from a forward-looking consumer with those of a more myopic consumer, we replicate the exercise of fitting moments from the end of Section 6.2. This time, when we fit the purchase hazard, purchase frequency, and interpurchase time variance for a consumer with a discount factor of 0.99, we allow $\omega_1$, $\omega_2$, and $\omega_3$ to vary, in addition to $\nu$ and $\pi_c$. As before, we are not able to fit the purchase hazard of a forward-looking consumer with that of a consumer with discount factor of 0.5. The fit appears to be somewhat better for higher values of $\nu$, however, for both values of $\nu$ the fit is especially poor in the early periods after a purchase occurs, precisely where our exclusion restrictions have the most identifying power. Intuitively, the exclusion restriction implies that consumers do not run out of packages very often, which means that the ability of storage costs to properly fit the purchase hazard is limited.

We note that technically all we need for identification is Rank Condition R2. Unless this condition fails, it will be impossible to exactly fit the purchase hazard of a forward-looking consumer with that of a myopic consumer. Our findings in this section provide some insight into what features of the data will help to identify forward-looking behavior, even in the presence of storage costs. In the following section, we will argue that in more realistic settings, we will observe even more variation in the data, which will make identification even more stark.

### 6.4 Price Variation and Purchasing Multiple Packages

Up to now we have limited our analysis to a somewhat hypothetical situation where prices are fixed over time, and consumers can only purchase a single package of a product. We have demonstrated that the discount factor can be identified even in this very limited situation. In many markets where consumers can stockpile, we observe periodic price promotions and consumers can purchase more than a single package of a product. This type of variation in the data can be used to make identification of the discount factor stronger. There are essentially two reasons for this. First, recall that when consumers are forward-looking, the expected future value of purchase increases as inventory drops. This implies that as inventory drops, a forward-looking consumer should become more sensitive to discounts. In contrast, in the absence of storage costs, a myopic consumer’s price sensitivity will remain constant until she runs out. This type of behavior will be observable in the sense that consumer sensitivity to prices should increase over the purchase hazard. To show this,
Figure 8: Probability of purchase in period $t$ given purchase of 1 package in period 0, for different values of $\omega_3$. Parameter values $\nu = 0.25, \pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \eta = 1, M = 3, p = 2$, and logit error term.

Figure 9: Fitted and actual purchase hazards for different true values of $\nu$, with true values of $\omega_1 = 0, \omega_2 = 0, \omega_3 = 0.5$. In the actual purchase hazard we set $\pi_c = 0.5$. Other parameter values $\eta = 1, M = 3, p = 2$, and logit error term.
we modify the model to allow for two different prices, which we set to 1 and 2. The transition probabilities between the different prices are Markov, where the probability of the price 2 given 2 occurred the previous period is 0.8, and the probability of 2 given last period’s price was 1 is 0.9. Thus most of the time prices are high, but periodically they drop to the low price for a short time, as is commonly observed in scanner data for storable goods. Figure 10 we plot the the difference between the probability of purchase at a low price versus a high price, as a function of the number of periods elapsed since a purchase of one package. The left panel shows the graph without storage costs, and the right graph shows with storage costs. As we described above, in both cases consumers who are forward-looking are more responsive to price cuts.

Figure 10: Probability of purchase at $p = 1$ minus probability of purchase at $p = 2$ in period $t$ given purchase of 1 package in period 0. Parameter values $\nu = 0.25, \omega_3 = 0.5, \beta = 0$, $\omega_3 = 0.5, \beta = 0.5$, $\omega_3 = 0.5, \beta = 0.75$, $\omega_3 = 0.5, \beta = 0.99$.

Consumers being able to purchase more than a single unit also helps identify the discount factor in the following sense: Without the influence of the error term, myopic consumers will never purchase more than a single package, which is all they need to avoid the stockout cost. However, forward-looking consumers may purchase more units if the price is sufficiently low. To show this, we do a similar exercise to the last paragraph, except instead of computing how the probability of purchase changes as time increases, we compute the difference between the ratio of the probability of buying 2 units to 1 unit. This ratio can be thought of as a consumer’s propensity to stockpile. For a forward-looking consumer, the propensity to stockpile should rise as prices drop, and as inventory drops. In Figure 11, we plot the ratio of probability of buying 2 units to 1 unit at the low price minus the same ratio at the high price, given $t$ periods have elapsed since the last purchase occurred. With no storage costs, this ratio rises over time for forward-looking consumers, but is
flat for myopic consumers. The intuition here is that the expected future payoff from stockpiling at a low price will rise as inventory drops. If there are storage costs, as shown in the right panel, then the probability of stockpiling in response to a low price may change over time even for myopic consumers, but the changes will be relatively small. Forward-looking consumers are less likely to stockpile initially, due to a potential increase in storage costs, but as they get closer to running out they will become more responsive to low prices.

Figure 11: Ratio of probability of buying 2 units to 1 unit at $p = 1$ minus the same ratio at $p = 2$ in period $t$, given purchase of 1 package in period 0. Parameter values $\nu = 0.25, \pi_c = 0.5, \omega_1 = 0, \omega_2 = 0, \omega_3 = 0.5, \eta = 1, M = 3$, and logit error term.

6.5 Identification with Consumer Unobserved Heterogeneity

All of the analysis above has assumed that there is no persistent unobserved heterogeneity across consumers. A formal argument for identification with persistent unobserved heterogeneity would rely on the time dimension of our data going to infinity at a rate that is fast enough relative to the cross-sectional dimension that one could estimate the purchase hazards and average purchase quantities described above on an individual basis. Using this type of argument one could, in principle, identify individual-specific discount factors. In field settings infinite amounts of data are not available, and so rather than allowing parameters to be individual specific, the researcher would have to rely on distributional assumptions about the unobserved heterogeneity to aid identification.
To provide further evidence that the model above can be identified in realistic settings, and to better understand when identification may become more difficult, we perform a series of artificial data experiments. As in the previous section, we perform our analysis on a dataset of 500 households who make purchases over 600 periods. We assume in period 1 everyone starts with 0 inventory; since in real data consumers will likely have been making purchases prior to the beginning of the data collection, we assume that the researcher only observes periods 201 to 600. The estimation method we use is in this section is simulated maximum likelihood. Since initial inventories are unobserved to the researcher, they must be simulated out. The approach we take is to use periods 201 to 400 to simulate initial inventories, and periods 401 to 601 to estimate parameters. In period 201 we assume all consumers begin with zero inventory, and draw a series of consumption shocks for each consumer. With simulated consumption shocks and observed purchase quantities one can construct an estimate of inventory in period 401. Our procedure of using the first part of a sample to construct inventories is standard in the literature (Erdem, Imai, and Keane (2003), Hendel and Nevo (2006a)). We use 100 simulated paths of consumption shocks for each household in constructing the likelihood.

To estimate the price coefficient we need sufficient price variation. We allow for 3 prices and use the price transition matrix shown in Table 2 to generate price processes. For the rest of the structure of the model, we allow consumers to purchase 2 packages at most, the package size $b = 8$, consumption shocks are in the set \{1, 2\}, and consumers can hold at most 3 packages. The error term is assumed to be logit and the weight on it is set to $\eta = 1$.

The results of the artificial data experiment are shown in Table 3. The top panel shows how the parameter identification is affected by including storage costs and by letting the storage cost function be more flexible. In the first 3 columns of this panel, we estimate the model in a situation where storage costs are zero. The first column shows the estimated parameters, the second the standard errors, and the third is the true values of the parameters. All the parameters are well

### Artificial Data Experiments

#### Table 2: Price Transitions Used in Artificial Data Experiment

<table>
<thead>
<tr>
<th>$p_{t-1}$</th>
<th>$p_t = 0.5$</th>
<th>$p_t = 1$</th>
<th>$p_t = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.1</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.1</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>
identified. The next three columns show how the results change if we allow $\omega_3$ to be free, while holding $\omega_1$ and $\omega_2$ fixed at 0. The parameter estimates are still close to the truth, although the standard errors are quite a bit larger. If we allow all 3 storage cost parameters to be positive, and estimate all of them, the standard errors rise significantly, although the parameter estimates are relatively close to the truth. We do not show the results when we allow for 2 storage costs to be free to save space; in that case the standard errors are a little higher than when we have only a single storage cost free. Increasing the number of storage cost parameters from 2 to 3 seems to increase the standard errors much more than increasing from 1 to 2. The fact that the precision on the discount factor drops as the number of storage cost parameters increases highlights the importance of exclusion restrictions. They allow for more precise identification of the discount factor when it is applies to more of the state space: for instance, in cases where it might be reasonable to assume that the cost of storing the first 1 or 2 packages is 0.

The bottom panel of the table shows how identification of the discount factor varies as consumers get more forward-looking. The first column shows the case where consumers are essentially myopic. In this case, the discount factor is not precisely identified. The reason for this is that a consumer with a positive, but low discount factor such as 0.3 behaves very similarly to a myopic consumer. As the discount factor rises, the precision with which we can estimate it also rises. In practice, the fact that the discount factor may not be precisely estimated for very myopic consumers is unlikely to be an issue: since the behavior of myopic consumers is similar over a wide range of low discount factors, counterfactual predictions over that range are likely to be similar as well.

Before turning to the empirical application, we note that we have also performed our artificial data experiments under the assumption that inventory is discrete rather than continuous. This exercise is presented in Appendix 12. Our findings in that section are similar: we can identify the discount factor well in general. We also experimented with trying to identify the discount factor under the assumption that the storage cost function was quadratic, and continuous in inventory, which is an assumption commonly made in empirical work. We find that even in this situation the discount factor is identified, although the standard errors are a bit higher than when the exclusion restriction applies. It is likely that the model’s functional form is driving identification in this case - technically, any sufficiently restrictive functional form assumption on the flow utility can allow the discount factor to be identified. The exclusion restriction approach is appealing in the sense that it is derived from institutional knowledge, rather than being an ad-hoc assumption about functional form.
Table 3: Artificial Data Experiment: Results

<table>
<thead>
<tr>
<th>Parameter</th>
<th>No Storage Costs</th>
<th>$\omega_2$ Free</th>
<th>$\omega_1, \omega_2$ Free</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>S.E.</td>
<td>Truth</td>
</tr>
<tr>
<td>Price Coeff ($\alpha$)</td>
<td>1.004</td>
<td>0.007</td>
<td>1</td>
</tr>
<tr>
<td>Stockout Cost ($\nu$)</td>
<td>0.098</td>
<td>0.01</td>
<td>0.1</td>
</tr>
<tr>
<td>Discount Factor ($\beta$)</td>
<td>0.957</td>
<td>0.016</td>
<td>0.95</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>$\omega_3$</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>$\pi_c$</td>
<td>0.489</td>
<td>0.007</td>
<td>0.5</td>
</tr>
</tbody>
</table>

8 Empirical Application

8.1 Data

To demonstrate how to apply our technique in practice 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. The data we are currently using covers the years 2001 through 2007. Estimation uses the final 3 years of the data while the first 4 are used to construct initial inventories. In our sample we include households who only purchase the 3 most popular sizes of detergent: the 100 oz size, the 128 oz size, and the 200 oz size. We restrict the sample to include households who purchase from the top 25 brands by overall purchase share. We also allow consumers to purchase multiple units of a size - for instance people will sometimes purchase 2 or 3 bottles of the 100 oz bottle. We remove households who ever purchase different products within the same week (this is very infrequent), or who purchase more than 3 bottles.
Table 4: Characteristics of Household Data

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of households</td>
<td>208</td>
</tr>
<tr>
<td>Avg interpurchase time (weeks)</td>
<td>17.1</td>
</tr>
<tr>
<td>Fraction of weeks with 0 units bought</td>
<td>0.952</td>
</tr>
<tr>
<td>Fraction of weeks with 1 unit bought</td>
<td>0.038</td>
</tr>
<tr>
<td>Fraction of weeks with 2 units bought</td>
<td>0.008</td>
</tr>
<tr>
<td>Fraction of weeks with 3 units bought</td>
<td>0.002</td>
</tr>
<tr>
<td>Fraction of purchases where 100 oz size chosen</td>
<td>0.702</td>
</tr>
<tr>
<td>Fraction of purchases where 128 oz size chosen</td>
<td>0.098</td>
</tr>
<tr>
<td>Fraction of purchases where 200 oz size chosen</td>
<td>0.2</td>
</tr>
</tbody>
</table>

of a product in a week. In this preliminary version of the paper we also work with a restricted sample. We only include households who make at least 5 purchases between 2005 and 2007, and for whom the maximum number of weeks between purchases is smaller than 40 weeks. This will cut out households who disappear from the sample for long periods of time, and who may be making laundry detergent purchases that aren’t recorded in the data. Our final sample contains 208 households.

Some statistics on our sample are shown in Table 4. An average household makes a purchase every 17 weeks, and in most weeks no purchase occurs. In our sample, consumers mostly purchase the smallest size bottle containing 100 ounces. Table 5 shows the purchase shares (the number bottles purchased of a particular brand divided by the total number of bottles purchased in the sample) as well as average prices (in cents per ounce) for each brand. (When constructing the sample we initially include the top 20 brands by purchase share. After reducing the sample to 208 households by removing those who purchase too infrequently or who purchase too much, only 15 brands have positive purchases)

8.2 Estimation Details

This section outlines the estimation procedure used to recover the model parameters from the IRI data. The main difference between the stylized model presented in Section 3 and the model we estimate is that the model we estimate contains different brands and different package sizes. Without simplifying assumptions, adding multiple brands and package sizes to the stylized model will complicate estimation considerably. These complications arise due to an increase in the size
Table 5: Brand Level Purchase Shares and Prices

<table>
<thead>
<tr>
<th>Brand</th>
<th>Purchase Share</th>
<th>Price (Cents Per Ounce)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tide</td>
<td>24.8</td>
<td>7.56</td>
</tr>
<tr>
<td>Xtra</td>
<td>9.4</td>
<td>2.53</td>
</tr>
<tr>
<td>Purex</td>
<td>10.3</td>
<td>3.93</td>
</tr>
<tr>
<td>All</td>
<td>10.3</td>
<td>5.17</td>
</tr>
<tr>
<td>Arm &amp; Hammer</td>
<td>9.9</td>
<td>4.69</td>
</tr>
<tr>
<td>Dynamo</td>
<td>5.4</td>
<td>5.14</td>
</tr>
<tr>
<td>Wisk</td>
<td>8.2</td>
<td>5.02</td>
</tr>
<tr>
<td>Era</td>
<td>13.6</td>
<td>6.17</td>
</tr>
<tr>
<td>Private Label</td>
<td>4</td>
<td>3.03</td>
</tr>
<tr>
<td>Fab</td>
<td>1.2</td>
<td>5.84</td>
</tr>
<tr>
<td>Yes</td>
<td>1.1</td>
<td>4.34</td>
</tr>
<tr>
<td>Ajax Fresh</td>
<td>0.3</td>
<td>3.22</td>
</tr>
<tr>
<td>Gain</td>
<td>0.4</td>
<td>6.38</td>
</tr>
<tr>
<td>Ajax</td>
<td>0.6</td>
<td>3.16</td>
</tr>
<tr>
<td>All Small &amp; Mighty</td>
<td>0.6</td>
<td>2.21</td>
</tr>
</tbody>
</table>

of the state space. Without changing the model, adding multiple brands implies we would have to track inventory for each brand, and a price for each brand separately. Additionally, adding multiple sizes means that a consumer’s state space would also need to track the number of bottles of each size held in inventory. The inventory composition would matter to the consumer since her storage cost will decrease as she uses up a bottle. A consumer who has two small bottles in her inventory will lower her storage cost more quickly than someone who has two large bottles. An additional complication is that multiple package sizes would require us to model the order in which packages are consumed. For instance, if a consumer has a large bottle and a small bottle in her inventory we would have to decide whether she would use the small bottle before the large one, or vice versa. Below we will first describe how we handle the issues arising from including different bottle sizes, and then from including different brands.

To deal with the issue related to multiple package sizes, we make a simplification in how we model storage costs. Specifically, we assume that each consumer has an upper bound on the amount that they can store, which we call \( \bar{\omega}_i \), and that storage costs are zero before the amount stored
hits \( \omega_i \). An intuitive explanation for this assumption is that consumers have some storage space dedicated to bottles of laundry detergent, and that they don’t purchase more bottles than what they can put in that space. We make an implicit assumption that the amount of space taken up by the bottles is proportional to the volume of the bottle (it might be the case that larger volume bottles are thinner and taller, and can be stored more easily than their volume would suggest - however, we have no data to verify this). Formally, our formulation of the storage cost is

\[
s(I, \omega_i) = \begin{cases} 
0 & \text{if } I \leq \omega_i \\
\infty & \text{otherwise}
\end{cases}
\]

Our assumptions solve the problems with the state space because storage costs do not depend on the number of bottles held (as long as they are below the bound). Thus, we do not have to model the composition of bottles in inventory or the order in which bottles are used. Even though we have put inventory directly into the storage cost function for convenience, the exclusion restriction still holds since storage costs are zero until the bound is reached, and a consumer will never purchase more than she can store. The storage cost bound \( \omega_i \) is an individual specific parameter we estimate.

To deal with issues arising from brand differentiation we follow Hendel and Nevo (2006a) and make two simplifying assumptions: the first is that consumers only care about brand differentiation at the time of purchase, and the second is inclusive value sufficiency. First we formalize the first of these two assumptions. Since consumers can purchase different bottle sizes, we index each bottle size by \( x = 1,\ldots,X \). We assume that utility from purchasing size \( x \) of brand \( k \) is equal to \( \xi_{ixk} \), where for each size \( x \) one of the \( \xi_{ixk} \) coefficients is normalized to zero. Our assumption means that the consumption utility, \( \gamma_i \), does not depend on the brand purchased (as we argued earlier, the parameter \( \gamma_i \) is not identified so we normalize \( \gamma_i = 0 \)). Additionally, it means that consumers only track the level of inventory, and not its composition. The consumer’s flow utility function from buying \( j > 0 \) units of size \( x \) of brand \( k \) can be written down as:

\[
u_{it}(k, x, j, I_{it}, \varepsilon_{ijit}, p_{it}, c_{it}; \theta_i) = \begin{cases} 
\xi_{ixk} - s(B_{i,t+1}(j, c_{it}); \omega_i) - \alpha_i p_{ixkt} j + \varepsilon_{ijxkt} & \text{if } I_{it} + b(x) j \geq c_{it} \\
\xi_{ixk} - \nu_i \frac{c_{it}}{c_{it}} - (I_{it} + b(x) j) - \alpha_i p_{ixkt} j + \varepsilon_{ijxkt} & \text{otherwise}
\end{cases}
\]

where \( b(x) \) is the number of ounces in a bottle of size \( x \). Before we write down the Bellman equation we need to clarify the elements of the consumer’s choice set. Consumers can either purchase nothing \( j = 0 \), or purchase \( j > 0 \) units of a single brand-size combination \((k, x)\). Denoting the feasible set of \((j, k, x)\) combinations as \( \mathcal{C} \), the consumer value function can be written as
\[ V_{it}(I_{it}, p_{it}) = \int_{c_{it}} E_{c_{it}} \max_{(j,k,x) \in C} \{ u_{it}(k, x, j, I_{it}, c_{it}; \theta_i) + \beta_i E_{p_{i,t+1} \mid p_{it}} V(I_{i,t+1}, p_{i,t+1}) \} dF(c_i), \]

where \( p_{it} \) is a vector of brand-size level prices. Our first assumption means we only track one inventory state variable; our second assumption of inclusive value sufficiency simplifies the state space further by reducing the price state to a single index. Before we show how inclusive value sufficiency simplifies the value function, we introduce a bit of additional notation related to the choice set. Denote \( C \) as the set of feasible \((j,x)\) combinations and \( C(j,x) \) as the set of feasible \( k \) values for \((j,x)\).\(^8\) To see how inclusive value sufficiency works note that we can write down the Bellman equation as

\[
\int_{c_{it}} \ln \left( \sum_{(j,x) \in C} \exp \left\{ u_{it}(k, x, j, I_{it}, c_{it}; \theta_i) + \beta_i E_{p_{i,t+1} \mid p_{it}} V(I_{i,t+1}, p_{i,t+1}) \right\} \right) dF(c_i)
\]

\[
= \int_{c_{it}} \ln \left( \sum_{(j,x) \in C} \sum_{k \in C(j,x)} \exp \left\{ \xi_{ixk} 1\{j > 0\} - \nu_i \frac{c_{it} - (I_{it} + b(x)j)}{c_{it}} 1\{I_{it} < c\} - \alpha_i p_{itzk} j \right\} \right) dF(c_i)
\]

\[
= \int_{c_{it}} \ln \left( \sum_{(j,x) \in C} \exp \left( \sum_{k \in C(j,x)} \exp \left( \xi_{ixk} 1\{j > 0\} - \alpha_i p_{itzk} j \right) \right) \right) - \nu_i \frac{c_{it} - (I_{it} + b(x)j)}{c_{it}} 1\{I_{it} < c\} + \beta_i E_{p_{i,t+1} \mid p_{it}} V(I_{i,t+1}, p_{i,t+1}) \right\} dF(c_i)
\]

\[
= \int_{c_{it}} \ln \left( \sum_{(j,x) \in C} \exp \left\{ \Omega_{it}(x,j) - \nu_i \frac{c_{it} - (I_{it} + b_j)}{c_{it}} 1\{I_{it} < c\} + \beta_i E_{p_{i,t+1} \mid p_{it}} V(I_{i,t+1}, p_{i,t+1}) \right\} \right) dF(c_i)
\]

Rather than tracking a \( K \) dimensional vector of prices, we assume that consumers track the index \( \Omega_{it}(x,j) \) for each possible size they could purchase. Alone, this simplification means that consumers track \(|C|\), rather than \(|\overline{C}|\) state variables.\(^9\) When we have solved our model numerically with inclusive value sufficiency we have found a further simplification we can make. We have found that we can explain about 99% of the variation in \( \Omega_{it}(x,j) \) for \( j > 1 \) with a spline regression of \( \Omega_{it}(x,j) \) on \( \Omega_{it}(x,1) \).\(^10\) This finding is not surprising, given that the functional forms of the \( \Omega \)

\(^8\)This is the empty set for \( j = 0 \), and can vary for different values of \( x \) since not all brands are available in all sizes.

\(^9\)|\(C| = \sum_{(j,x) \in C} = 1 + (J - 1)X, \) and \(|\overline{C}|\) will be \( \sum_{(j,x) \in C} \sum_{k \in C(j,x)} > |C|\). If all brands were available for all sizes then the dimensionality reduction would be \( K \).

\(^10\)Note that \( \Omega_{it}(0) \) is 0, since consumers get no brand utility if they do not buy anything.
indexes are so similar. Rather than assuming that consumers track a separate index for each size, we assume that they only track \( \Omega_{it}(x,1) \), and construct the predicted value of \( \Omega_{it}(x,j) \) from the spline regression. We will denote the fitted \( \Omega_{it}(x,j) \) as \( \hat{\Omega}_{it}(x,j) \).\(^{11}\) The state space of the inclusive values is of dimensionality \( X \), and we will write the inclusive value state variable as \( \Omega_{it}(1) \). Under these assumptions the consumer Bellman equation can be written as

\[
V_{it}(I_{it}, \Omega_{it}(1)) = \int_{c_{it}} \ln \left( \sum_{j=0}^{J} \exp\{\hat{\Omega}_{it}(x,j) - \nu_i \frac{c_{it} - (I_{it} + b(x)j)}{c_{it}} \} 1\{I_{it} < c\} + \beta_i E_{\Omega_{i,t+1}(1)} V(I_{i,t+1}, \Omega_{i,t+1}(1)) \right) dF(c_{it}).
\]

To estimate the model we use the Bayesian estimation method of Imai, Jain, and Ching (2009) (henceforth abbreviated IJC). Our implementation of Imai, Jain, and Ching (2009) also builds on the work of Hendel and Nevo (2006a). However since our approach is Bayesian, it differs from Hendel and Nevo (2006a), propose a 3 step method that uses maximum likelihood. In the first step one estimates the price coefficient \( \alpha \) and the brand specific coefficients \( \xi_{ixk} \). Then, one constructs the inclusive values \( \Omega_{it}(1) \) and estimates the transition process for them. Finally, one estimates the remaining model parameters (such as the stockout cost, consumption rates, and in our case the discount factor) using maximum likelihood. The likelihood is formed over a consumer’s quantity choice. The reason we choose to use IJC’s Bayesian approach is computational. We found that the third step of Hendel and Nevo (2006a) was computationally burdensome in our setting due to the fact that one must continuously solve the value function as one optimizes the likelihood over quantity choices. In our case the size of the state space is 4 dimensional. We found that when we solved for the value function numerically, to get acceptable accuracy it was necessary to use 10 points for each dimension of the inclusive value and 100 for the inventory dimension, which means we have to solve the value function on a grid of 100,000 points. We found it took several minutes to solve for the value function on this grid using value function iteration. Note that this was done for a single, fixed parameter draw. We believe it is important to model unobserved heterogeneity in consumer preferences, which would substantially increase the computational burden of the third step.

The IJC method has the advantage that one does not have to solve the value function repeatedly - rather one iterates on the value function over the course of the MCMC chain making solution much faster. Additionally it can more easily incorporate unobserved heterogeneity across consumers.

\(^{11}\)For notational convenience we will use \( \hat{\Omega}_{it}(x,1) \) and \( \Omega_{it}(x,1) \) interchangeably.
In addition to the different specification used for storage cost, we make three other more minor changes to the model specification from the specification used for simulation. First, we incorporate a carrying cost $CC_i$, which is the disutility a consumer receives from making a purchase. Second, we assume that an individual’s consumption rate is a fixed parameter so that $c_i = \tau_i = c_i$. We plan to relax this latter assumption in future revisions to the paper, however we believe the assumption of a fixed consumption rate is not unreasonable for a product category like laundry detergent where consumers probably do laundry at regular intervals. The third change is that we find occasionally there are some weeks where consumers do not visit any store. To capture this we assume that there is an exogenous probability a consumer goes to the store, which we estimate prior to estimating the other model parameters. This probability is incorporated into consumers’ expectations when they update their value functions. For simplicity of our exposition below, we outline how the solution of the model works when it is assumed that consumers always visit a store. We allow for unobserved heterogeneity in all model parameters except for two of the brand coefficients - we found we could not identify all the variances of all the brand coefficients. The basic steps of the algorithm are as follows:

1. Draw population-varying parameters using Metropolis-Hastings
2. Draw mean of population-varying parameters
3. Draw variance of population-varying parameters
4. Draw population-fixed parameters using Metropolis-Hastings
5. Update value function

We describe how we implement steps 1 to 4 in Section 8.3, and step 5 in Section 8.4. Some other details related to the construction of the inclusive value transition process and setup of the MCMC chain are described in Sections 8.5 and 8.6. Readers who are not interested in the implementation details of our estimation methodology may skip to the estimation results in Section 8.7. Before continuing with the estimation details we introduce some additional notation. Denote the vector of population-varying parameters drawn in step 1 as $\theta_{i1}$, and the population-fixed parameters in step 2 as $\theta_2$. We assume that the individual- specific parameters are derived from a normal distribution with mean $b$ and variance $W$. Since some of the parameters must be bounded (such as the discount factor or price coefficient) we assume that they are transformations of underlying normal parameters. We assume that the price coefficient, the stockout cost, and the consumption rates are lognormal. The transformation applied to produce the discount factor is $\exp(x)/(1 + \exp(x))$. 

42
where \( x \) is normal. The inventory bound transformation is \( 600 \times \exp(x)/(1 + \exp(x)) \). We bound the inventory bound parameter at 600 because we evaluate the value function on a grid of points where the maximum is 600 ounces. This bound on the state space implies that nobody can hold more than 3 of the largest size of bottles in their inventory. We will denote the untransformed parameters as \( \tilde{\theta}_{i1} \), and the transformed parameters as \( \theta_{i1} = T(\tilde{\theta}_{i1}) \). Note that we assume that \( \tilde{\theta}_{i1} \sim N(\mathbf{b}, \mathbf{W}) \).

**8.3 Steps 1 to 4: Drawing the model parameters**

We use the random walk Metropolis-Hastings Algorithm to implement Step 1 of the Gibbs sampler, and draw the individual specific parameters on a household-by-household basis. To that end we describe how we draw an individual \( \theta_{i1} \). Suppose that we are at step \( g \) of the Gibbs sampler. First, conditional on the last step’s draw of \( \tilde{\theta}_{i1} \), which we call \( \tilde{\theta}_{i1}^{0} \), we draw a candidate \( \tilde{\theta}_{i1}^{1} \) from \( N(\tilde{\theta}_{i1}^{0}, \rho_{1}W_{g-1}) \), where \( W_{g-1} \) is last iteration’s estimate of the variance matrix. Our new utility parameters will be \( \theta_{i1}^{1} = T(\tilde{\theta}_{i1}^{1}) \). We then compute the joint likelihood of brand and size purchase at the old draw and the candidate draw. To implement this we first need an estimate of each consumer’s value function. As we describe further in Section 8.4, we compute this estimate by averaging over past value functions, using the nearest neighbor approach of Norets (2009). The choice probability can be written as the probability of the observed brand choice \( (k_{it}) \) given size choice \( (x_{it}, j_{it}) \), multiplied by the probability of the observed size choice. For a given individual the probability of a particular brand choice given their size choice is

\[
Pr(k_{it}|x_{it}, j_{it}, p_{it}; \theta_{i1}, \theta_{2}) = \frac{\exp(\xi_{x_{it}, k_{it}} - \alpha p_{i, x_{it}, k_{it}, t} \star j_{it})}{\sum_{l \in C} \exp(\xi_{x_{it}, l} - \alpha p_{i, x_{it}, l, t} \star j_{it})}. \tag{12}
\]

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

\[
Pr(x_{it}, j_{it}|\Omega_{it}(1); \theta_{i1}, \theta_{2}) = \frac{\exp \left( \hat{\Omega}_{it}(x_{it}, j_{it}) + \hat{u}(I_{it}, j_{it}, x_{it}; \theta_{i1}, \theta_{2}) + \beta \hat{EV}_{i}(I_{i,t+1}, \Omega_{it}(1); \theta_{i1}, \theta_{2}) \right)}{\sum_{(j, x) \in C} \exp \left( \hat{\Omega}_{it}(x, j) + \hat{u}(I_{it}, j, x; \theta_{i1}, \theta_{2}) + \beta \hat{EV}_{i}(I_{i,t+1}, \Omega_{it}(1); \theta_{i1}, \theta_{2}) \right)}, \tag{13}
\]

where \( \hat{EV}(I_{i,t+1}, \Omega_{it}(1); \theta_{i1}, \theta_{2}) \) is the estimated expected value function, and

\[
\hat{u}(I_{it}, j_{it}, x_{it}; \theta_{i1}, \theta_{2}) = -\nu_{t} \frac{c_{it} - (I_{it} + bj)}{c_{it}} \mathbb{1}(I_{it} < c).
\]

Note that to compute this probability we need to compute the inclusive values \( \Omega \), which themselves are functions of \( \theta_{i1} \) and \( \theta_{2} \) parameter draws. To construct the estimated value function we will
also need to compute the transition process for the inclusive values. We discuss how the inclusive values and their transition process are computed in Section 8.5.

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

\[ L_i(\theta_{i1}, \theta_2) = \prod_{t=1}^{T_i} Pr(k_{it}|x_{it}, j_{it}, p_{it}; \theta_{i1}, \theta_2) Pr(x_{it}, j_{it}|\Omega_{it}(1); \theta_{i1}, \theta_2). \]

The candidate draw will be accepted with probability

\[ \frac{L(\theta_{i1}^1, \theta_2) k(\theta_{i1}^0)}{L(\theta_{i1}^0, \theta_2) k(\theta_{i1}^1)}, \]

where \( k \) denotes the prior density on \( \theta_{i1} \). Under our assumption of normality of the parameters this prior is simply the multivariate normal with mean \( b_{g-1} \) and variance \( W_{g-1} \).

After drawing the population-varying parameters we draw the mean \( (b) \) and variance \( (W) \) parameters that generate them (Steps 2 and 3). We assume a diffuse prior on \( b \), so the next draw is taken from a normal with mean \( \bar{\theta}_{i1,g} \), and variance \( W_{g-1}/N \), where \( N \) is the number of households. We also assume a diffuse prior on \( W \). If the dimensionality of \( \bar{\theta}_{i1,g} \) is \( K \) then the posterior on \( W \) is inverted Wishart with \( K + N \) degrees of freedom and a scale matrix of \((KI + NS_1)/(K + N)\)

where \( S_1 = 1/N \sum_{i=1}^{N} (\bar{\theta}_{i1,g} - b)(\bar{\theta}_{i1,g} - b)' \).

The fourth step is to draw the population-fixed parameters \( \theta_2 \). This step proceeds in largely the same way as the first step. A candidate draw \( \theta_{i2}^1 \) is taken from \( N(\theta_{i2}^0, \rho_2 W_2) \), and is accepted with probability

\[ \frac{\prod_{i=1}^{I} L_i(\theta_{i1}, \theta_{i2}^1) k(\theta_{i2}^0)}{\prod_{i=1}^{I} L_i(\theta_{i1}, \theta_{i2}^0) k(\theta_{i2}^1)}. \]

We assume a diffuse prior on \( \theta_2 \).

### 8.4 Step 5: Updating the value function

After computing the current step draws the value functions are updated at the current draw. For each individual in the data, we store the value function on a grid of 100 inventory points and 100 random inclusive value draws (meaning we update the value function for each consumer on a grid of 10,000 points). At a particular grid point, we need to compute an estimate of the value function at the current draw. Denote this value function estimate as \( \hat{EV}_i(I, \Omega; \theta_{i1}, \theta_2) \). We will index inventory grid points (which are fixed across Gibbs iterations) with \( s_1 \) and inclusive value grid points (which are random) with \( s_2 \). Then at a particular grid point \( s_1, s_2 \), with state variables \( I_{s_1}, \Omega_{s_2} \), the updated value function is
\[ \hat{V}_i(I_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) = \sum_{(j,x) \in C} \log \left( \exp \left( \hat{\Omega}_{s_2}(x, j) + \hat{u}(I_{s_1}, j, x; \theta_{i1}, \theta_2) + \beta F V_i(I'_{s_1}, \Omega_{s_1}; \theta_{i1}, \theta_2) \right) \right). \]  

(14)

There are two issues we face when computing the expected value function \( E \hat{V}(I'_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) \) in equation (14). First, we need to do it at a new parameter draw, \( \theta_{i1}, \theta_2, \) for which we have no prior estimate of the value function; second, we need to integrate out future values of the inclusive value. We address the first issue by averaging over past value functions at parameters draws close to the current draw of \( \theta_{i1}, \theta_2, \) following Imai, Jain, and Ching (2009). We address the second by using a random grid approach combined with importance sampling following Norets (2009). In each Gibbs iteration we draw \( N_s = 100 \) inclusive value draws \( \Omega_{s_2} \) from an importance distribution \( h(\cdot) \). We choose a multivariate normal for the importance distribution and estimate its moments prior to running the MCMC chain. To do this we estimate brand and choice parameters by running a multinomial logit of brand choice conditional on size, and then use those parameters to construct preliminary estimates of the inclusive value for each package size at each data point. The mean of \( h(\cdot) \) is the average estimated inclusive value for each size and the variance is the variance matrix of the sizes across all observations. We denote the transition density of the \( \Omega \)'s as \( F(\Omega_t | \Omega_{t-1}; \theta_{i1}, \theta_2) \), and discuss how we construct this density in the next section.

Turning to the details of the procedure, for the past \( G \) Gibbs draws we have saved estimates of the value function \( \hat{V}_i \) at each of the inventory states \( s_1 \) and the past importance draws on the inclusive values \( s_2 \). Denote the saved value functions and draws as \( \hat{V}_i^g, \theta_{i1}^g, \theta_2^g, \quad g = 1, \ldots, G \). Our first step is to find the indexes of the \( N(G) \) saved draws \( \theta_{i1}^g, \theta_2^g \) that are closest to \( \theta_{i1}, \theta_2 \). Denote these indexes as \( S(G) \). We then compute an intermediate estimate of the value function we denote \( \tilde{V}_i(I_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) \) for each of the inventory and inclusive value grid points using importance sampling:

\[ \tilde{V}_i(I_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) = \sum_{g \in S(G)} \sum_{n_s=1}^{N_s} \frac{\hat{V}_i^g(I_{s_1}, \Omega_{s_2}; \theta_{i1}^g, \theta_2^g) F(\Omega_{s_2}^g | \Omega_{s_1}; \theta_{i1}^g, \theta_2^g)}{h(\Omega_{s_2}^g)} . \]

When we compute the expected value function, \( E \hat{V}(I'_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) \), we need to account for the fact that the next period’s inventory \( I'_{s_1} \) may not coincide with the grid points for inventory indexed by \( s_1 \). To address this we interpolate the \( \tilde{V}_i(I_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) \) over inventory states using linear interpolation. Then the expected value function estimate is

\[ E \hat{V}_i(I'_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) = \tilde{V}_i(I'_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) + \frac{I'_{s_1} - I_{s_1}}{I'_{s_1} - I_{s_1}} \tilde{V}_i(I'_{s_1}, \Omega_{s_2}; \theta_{i1}, \theta_2) . \]
where $I_{s1}'$ is the largest inventory grid point that is smaller than $I_{s1}'$ and $T_{s1}'$ is the smallest grid point that is larger than $I_{s1}'$.

Once we have computed $\hat{V}_i(I_{s1}, \Omega_{s2}; \theta_1; \theta_2)$ at the current draw for all state space points we store it in memory for use in future value function approximations. Note that when we compute choice probabilities in Section 8.3 we also need to compute the expected value function; for this we use a similar procedure to the above.

### 8.5 Inclusive value transition process

When we take a new draw on the parameters $\theta_1$ and $\theta_2$ we need to compute new inclusive values, as well as to estimate their transition processes, and the functions for approximating the inclusive values for $j > 1$. First we compute the inclusive values $\Omega_{it}(x, j)$ for each household and time period.

Then we compute a B-Spline basis for $\Omega(x, 1)$ for each value of $x$ and each individual $i$ in the data (Habermann and Kindermann 2007). To compute the basis we first compute the minimum and maximum values of $\Omega_{it}(x, 1)$ for each value of $x$ and each individual $i$, and decide on a number of equidistant interpolation nodes $n_d$. Denoting the minimum and maximum $\Omega_{it}(x, 1)$ as $a_{i,x}$ and $b_{i,x}$ respectively, the nodes are defined as $\Omega_{k,i}(x, 1) = a_{i,x} + kh_i$, $h_i = (b_{i,x} - a_{i,x})/n_d$, $k = 0, \ldots, n_d$.

For a particular value of $\Omega(x, 1)$ the spline basis is the vector of functions

$$ u_{ik}(\Omega) = \Phi \left( \frac{\Omega - a_{i,x}}{h_i} - (k - 2) \right), \quad k = 1, \ldots, n_d + 3, $$

where

$$ \Phi(t) = \begin{cases} 
(2 - |t|)^3 & 1 \leq |t| \leq 2 \\
4 - 6|t|^2 + 3|t|^3 & |t| \leq 1 \\
0, & \text{elsewhere}
\end{cases} $$

Our approach for constructing $\hat{\Omega}(x, j)$ for $j > 1$ is to use the predicted value of the regression

$$ \Omega_{it}(x, j) = \sum_{k=1}^{n_d+3} \kappa_{i,x,k} u_{ik}(\Omega_{it}(x, 1)) + \epsilon_{it,x}, \quad (15) $$

where the $\kappa_{i,x,k}$ coefficients are estimated using OLS.

We also use a spline basis approach to model the autocorrelation process for $\Omega(1)$. We estimate the regression equation

$$ \Omega_{i,t+1}(1) = u_i(\Omega_{i,t}(1))' \psi_i + \epsilon, \quad (16) $$

where $u_i(\Omega_{i,t}(1))$ is a $X \times (n_d + 3)$ matrix of spline bases. Note that in this regression we regress each $\Omega_{i,t+1}(x, 1)$ on the $u_k$ spline basis functions for all the previous period $\Omega_{i,t+1}(x, 1)$'s for
all the values of $X$, since prices across different sizes are correlated. We also place no restrictions on the error variance matrix of $\epsilon$. Note that the parameters of the inclusive value regressions are individual specific since they depend on the individual specific $\theta_{it}$ and individual $i$’s data.\footnote{We found when estimating the inclusive value regression for sizes greater than 1 in equation (15), not all the coefficients could always be identified, since some sizes of some products were not always carried much in stores where the individual shopped. To address this problem we estimate these coefficients for each individual on a random sample of prices. We take 10% of the observations in our data where a store is visited.} Each regression is run separately for each individual $i$.

### 8.6 Setup of the Gibbs Sampler

In this section we describe some details of the setup of the Gibbs sampler. The computer code we use is written in R and C and designed to take advantage of parallel processing in the value function averaging and updating. Our R code for Bayesian estimation makes use of routines from Rossi, Allenby, and McCulloch (2005) for summarizing the model output. For the Metropolis-Hastings steps in steps 1 and 4 of the Gibbs sampler we need to set the parameters $\rho_1$ and $\rho_2$, which are both tuned so that the acceptance rate is about 30%. We tune the parameter $\rho_1$ every iteration: if the fraction of household level parameters that are accepted is above 30%, we increase $\rho_1$ by 10%; otherwise we decrease it by 10%. For $\rho_2$, we adjust the parameter every 25 iterations: if the number of acceptances for the past 25 iterations is above 30%, then the $\rho_2$ parameter is decreased by 25%; otherwise it is increased by 25%. The $\rho$ parameters move some initially but settle down after about 500 iterations.

For the value function approximation we choose $G = 10$ and $N(G) = 3$ (Norets (2009) finds in simulations that setting $G = 1$ and $N(G) = 1$ is sufficient to get a reasonable approximation to the value function since the Gibbs draws are autocorrelated - our choices for these variables are thus conservative). We use standard Euclidean distance to compute the closeness of the $\theta_{1i}$ and $\theta_{2i}$ parameter vector to past draws. As we discussed above we evaluate the value function on 100 grid points. Grid points for inventory are chosen between 0 and the maximum inventory value of 600 ounces, with 50% of the points being clustered equally between 0 and 120 and the rest between 120 and 600. We choose more points near zero since the value function tends to be more nonlinear in that region, and we want a better approximation there. Note that these grid points are fixed for all the estimation draws. For the B-Spline basis function we choose the number of nodes $n_d = 3$. This number of nodes results in very good fitted values ($R^2$’s of around 99% on our regressions). We assume that the error process for the inclusive value autoregression error $\epsilon$ is multivariate normal.

We run the Gibbs sampler for 10,000 iterations. The draws appear to converge at about 2,000
iterations, so we drop the first 2,000 draws to reduce burn-in.

8.7 Estimation Results

This section presents our estimation results. Since most of the parameters vary across the population, we present the averages of some of the population moments of each parameter: the lower 25th percentile, the median, the mean, and the 75th percentile. Table 6 shows the estimates of the brand parameters. The first row shows the estimated population distribution of tastes for Xtra 200 oz. To compute an estimated moment, say the 25th percentile, first for each Gibbs draw we compute the 25th percentile of the population distribution of taste draws for Xtra. The estimated 25th percentile is the average of the 25th percentiles over all 8,000 saved Gibbs draws.\(^{13}\) The second row shows the 95% confidence bounds on each of the estimated moments. The overall takeaway from this table is that there is a significant amount of heterogeneity in tastes for products, which is not surprising. Table 7 shows the estimates of the other model parameters. The consumption rate parameters suggest a significant amount of heterogeneity across the population, and indicate that the median household takes about 4 weeks to finish the smallest sized bottle of detergent. There is a significant amount of unobserved heterogeneity in the price coefficient, as well as the stockout cost. The large negative carrying costs rationalize the low frequency of purchase. The stockout cost estimates imply that most consumers can store 1 or 2 of the largest, 200 ounce size bottles of detergent.

The fourth row of Table 7 shows the estimated population distribution of the discount factor, which is the focus of the paper. Our results suggest that most consumers behave in a less forward-looking way than most papers which estimate forward-looking models of consumer behavior assume. Papers which assume a discount factor typically set consumer discount factors to a value such as 0.95, 0.99 or 0.999. In contrast we find that the median consumer’s discount factor is 0.93, while the average is 0.95. Very few consumers have discount factors on the order 0.99 - our estimates imply that this is the upper 1% tail of the population. To get a sense of the population distribution of discount factors, in Figure 12 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). Most individuals' discount factors lie between 0.85 and 0.95. Although our estimated discount factors are less than what past work has assumed, the fact that they are lower than what is typically assumed is consistent the field study of Yao, Mela, Chiang, and Chen (2012), who estimate in data on cellular phone usage that consumer discount factors are around 0.91.

\(^{13}\)In Figure 13 of Section 10 we plot the estimated mean parameter \(b\) at each of the 10,000 Gibbs steps for selected parameters. The parameters seem to stabilize at or before draw 2,000.
Taking our results at face value it may be tempting to argue that our estimates suggest consumers are irrational, as weekly a discount factor of 0.91 would translate to a yearly discount factor of close to 0, implying consumers are essentially myopic when making financial decisions where the time horizon was on the order of a year. However, there may be other possible explanations that are more consistent with rational behavior. One possibility is that consumers behave in a less forward-looking way when they make decisions about purchasing consumer products due to the difficulty involved in figuring out the optimal purchase rule. When making important financial decisions consumers may behave in a more forward-looking way due to the fact that more money is on the line - there are more gains to making the right decision.

![Discount Factor Estimate](image)

Figure 12: Kernel Density of Individual-Specific Discount Factor Estimates.

9 Conclusion

Consumer stockpiling behavior in consumer package goods is often cited as an example where consumers are forward-looking. However, previous research (most notably, Erdem, Imai, and Keane (2003), Hendel and Nevo (2006a)) assumes (i) consumer are homogeneous in their discount factors, and (ii) consumers do not arbitrage and hence discount factor can be set according to the prevailing interest rate. By exploring exclusion restrictions that have not been previously studied, and using recently developed estimation methods, we are able to relax these two restrictions. To classical economists, our findings may be surprising because consumers are not only heterogeneous in their discount factors, but their magnitudes are also significantly lower than what the interest rate
<table>
<thead>
<tr>
<th>Parameter</th>
<th>1st Quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xtra, 200 oz size</td>
<td>-7.03</td>
<td>-6.4</td>
<td>-6.38</td>
<td>-5.75</td>
</tr>
<tr>
<td>Purex, 100 oz size</td>
<td>-2.16</td>
<td>-1.51</td>
<td>-1.49</td>
<td>-0.86</td>
</tr>
<tr>
<td></td>
<td>[-2.46, -1.89]</td>
<td>[-1.8, -1.24]</td>
<td>[-1.77, -1.28]</td>
<td>[-1.14, -0.6]</td>
</tr>
<tr>
<td>Purex, 128 oz size</td>
<td>-5.25</td>
<td>-4.61</td>
<td>-4.61</td>
<td>-3.97</td>
</tr>
<tr>
<td>Purex, 200 oz size</td>
<td>-4.68</td>
<td>-4.05</td>
<td>-4.04</td>
<td>-3.4</td>
</tr>
<tr>
<td>All, 100 oz size</td>
<td>-1.61</td>
<td>-0.99</td>
<td>-0.92</td>
<td>-0.32</td>
</tr>
<tr>
<td></td>
<td>[-1.83, -1.39]</td>
<td>[-1.23, -0.76]</td>
<td>[-1.14, -0.75]</td>
<td>[-0.57, -0.1]</td>
</tr>
<tr>
<td>All, 200 oz size</td>
<td>-4.58</td>
<td>-3.93</td>
<td>-3.93</td>
<td>-3.29</td>
</tr>
<tr>
<td>Arm &amp; Hammer, 100 oz size</td>
<td>-2.02</td>
<td>-1.38</td>
<td>-1.35</td>
<td>-0.72</td>
</tr>
<tr>
<td></td>
<td>[-2.37, -1.76]</td>
<td>[-1.71, -1.12]</td>
<td>[-1.67, -1.11]</td>
<td>[-1.04, -0.45]</td>
</tr>
<tr>
<td>Arm &amp; Hammer, 200 oz size</td>
<td>-7.09</td>
<td>-6.44</td>
<td>-6.44</td>
<td>-5.8</td>
</tr>
<tr>
<td>Dynamo, 100 oz size</td>
<td>-2.89</td>
<td>-2.27</td>
<td>-2.22</td>
<td>-1.62</td>
</tr>
<tr>
<td></td>
<td>[-3.53, -2.53]</td>
<td>[-2.87, -1.88]</td>
<td>[-2.85, -1.86]</td>
<td>[-2.24, -1.23]</td>
</tr>
<tr>
<td>Dynamo, 200 oz size</td>
<td>-4.64</td>
<td>-4</td>
<td>-3.99</td>
<td>-3.35</td>
</tr>
<tr>
<td>Wisk, 100 oz size</td>
<td>-2.68</td>
<td>-2.06</td>
<td>-2.04</td>
<td>-1.41</td>
</tr>
<tr>
<td></td>
<td>[-3.19, -2.3]</td>
<td>[-2.59, -1.69]</td>
<td>[-2.58, -1.7]</td>
<td>[-1.98, -1.05]</td>
</tr>
<tr>
<td>Era, 100 oz size</td>
<td>-1.21</td>
<td>-0.61</td>
<td>-0.49</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>[-1.47, -0.95]</td>
<td>[-0.87, -0.35]</td>
<td>[-0.73, -0.26]</td>
<td>[-0.21, 0.34]</td>
</tr>
<tr>
<td>Private Label, 100 oz size</td>
<td>-5.14</td>
<td>-4.5</td>
<td>4</td>
<td>-3.86</td>
</tr>
<tr>
<td></td>
<td>[-5.9, -4.43]</td>
<td>[-5.26, -3.8]</td>
<td>[-5.22, -3.83]</td>
<td>[-4.57, -3.16]</td>
</tr>
<tr>
<td>Private Label, 128 oz size</td>
<td>-3.5</td>
<td>-2.87</td>
<td>-2.85</td>
<td>-2.23</td>
</tr>
<tr>
<td></td>
<td>[-3.9, -3.15]</td>
<td>[-3.27, -2.52]</td>
<td>[-3.24, -2.53]</td>
<td>[-2.65, -1.9]</td>
</tr>
<tr>
<td>Fab, 100 oz size</td>
<td>-3.91</td>
<td>-3.28</td>
<td>-3.27</td>
<td>-2.65</td>
</tr>
<tr>
<td>Yes, 100 oz size</td>
<td>-5.83</td>
<td>-5.19</td>
<td>-5.19</td>
<td>-4.55</td>
</tr>
<tr>
<td></td>
<td>[-6.27, -5.37]</td>
<td>[-5.62, -4.75]</td>
<td>[-5.61, -4.76]</td>
<td>[-5, -4.09]</td>
</tr>
<tr>
<td>Ajax Fresh, 128 oz size</td>
<td>-3.98</td>
<td>-3.34</td>
<td>-3.35</td>
<td>-2.71</td>
</tr>
<tr>
<td>Gain, 100 oz size</td>
<td>-5.08</td>
<td>-4.43</td>
<td>-4.43</td>
<td>-3.79</td>
</tr>
<tr>
<td></td>
<td>[-5.42, -4.46]</td>
<td>[-4.73, -3.83]</td>
<td>[-4.73, -3.84]</td>
<td>[-4.14, -3.2]</td>
</tr>
<tr>
<td>Ajax, 128 oz size</td>
<td>-3.6</td>
<td>-3.6</td>
<td>3.6</td>
<td>3.6</td>
</tr>
<tr>
<td>All Small &amp; Mighty, 128 oz size</td>
<td>-2.93</td>
<td>-2.93</td>
<td>-2.93</td>
<td>-2.93</td>
</tr>
</tbody>
</table>

Notes: This table shows average moments of the posterior distribution of the population distribution of the brand parameters. For example, the median columns shows the average of the population median of a given taste parameter, where the average is taken across MCMC draws. Square brackets show 95% confidence intervals. Brand coefficients for Tide 100oz, Tide 200oz and Xtra 128oz are normalized to zero. The final two products are normalized to be fixed across the population.
Table 7: Dynamic Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>1st Quartile</th>
<th>Median</th>
<th>Mean</th>
<th>3rd Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumption Rate</td>
<td>15.42</td>
<td>27.43</td>
<td>41.37</td>
<td>48.49</td>
</tr>
<tr>
<td></td>
<td>[10.92, 19.62]</td>
<td>[18.94, 33.6]</td>
<td>[27.87, 50.84]</td>
<td>[34.62, 60.59]</td>
</tr>
<tr>
<td>Price Coefficient</td>
<td>-0.44</td>
<td>-0.22</td>
<td>-0.34</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>[-0.52, -0.37]</td>
<td>[-0.26, -0.18]</td>
<td>[-0.38, -0.29]</td>
<td>[-0.14, -0.09]</td>
</tr>
<tr>
<td>Stockout Cost</td>
<td>0.41</td>
<td>0.75</td>
<td>1.2</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td>[0.23, 0.54]</td>
<td>[0.42, 0.99]</td>
<td>[0.65, 1.65]</td>
<td>[0.77, 1.82]</td>
</tr>
<tr>
<td>Discount Factor</td>
<td>0.88</td>
<td>0.93</td>
<td>0.91</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>[0.85, 0.91]</td>
<td>[0.92, 0.95]</td>
<td>[0.89, 0.93]</td>
<td>[0.95, 0.97]</td>
</tr>
<tr>
<td>Carrying Cost</td>
<td>-5.65</td>
<td>-4.51</td>
<td>-5.19</td>
<td>-3.4</td>
</tr>
<tr>
<td>Carrying Capacity</td>
<td>192.28</td>
<td>302.56</td>
<td>444.18</td>
<td>553.87</td>
</tr>
<tr>
<td></td>
<td>[176.11, 212.24]</td>
<td>[253, 365.51]</td>
<td>[376.14, 547.44]</td>
<td>[443.45, 715.88]</td>
</tr>
</tbody>
</table>

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

predicts. Our estimated weekly discount factors average at around 0.91, lower than the value of 0.99 this is obtained if one uses a common interest rate to set it. The differences are large and they could lead to material impact on the results of counterfactual experiments conducted in prior research which fixes consumer discount factors.

As an example, Hendel and Nevo (2006a) argue that the price coefficient in consumers’ utility function will be biased upwards if one ignores their forward-looking incentive (i.e., by setting the discount factor to be zero). As a result, the short-run price elasticities estimated from a static demand model can overestimate the dynamic ones by about 30 percent. Based on our current results, we expect that a static model would still overestimate the price coefficient in comparison to a dynamic one, but by a much smaller order of magnitude. For some consumers, such overestimation could be sufficiently small that a static demand model would do a fine job in measuring their price elasticities.

Hendel and Nevo (2006a) also argue that ignoring consumers’ forward-looking incentives would lead to overestimation of price-cost margins, underestimation of cross-price elasticities, and overes-
timates of the amount of substitution to the outside alternative. The latter two findings imply that using estimates from a static demand model could lead to misleading policy decisions in approving mergers (an antitrust authority that relied on static demand estimates would be too lenient). However, our findings suggest that the standard practice of setting the discount factor using the prevailing interest rate could generate the opposite outcome, i.e., price-cost margins would be underestimated, and hence merger decisions would be made too conservatively. In particular, it is common for researchers to use price-cost margins to test whether firms collude. Using a discount factor that is too high would increase the incidence of type I errors, i.e., rejecting the collusion hypothesis when it is indeed happening.

References


Figure 13: Appendix Figure: Plots of Gibbs Draws for Selected Dynamic Parameters
11 Appendix: Proofs

11.1 Proof of Lemma 1

To begin, we propose an optimal policy which is that consumers only purchase when they run out. Under this optimal policy the value functions will be:

\[ V(0) = -p + \pi_c \beta V(b - 1) + (1 - \pi_c) \beta V(b - 2) - \omega_1, \]

\[ V(1) = \pi_c \beta V(0) + (1 - \pi_c)(-p + \beta V(b - 1) - \omega_1), \]

\[ V(2) = \pi_c (\beta V(1) - \omega_1) + (1 - \pi_c) \beta V(0), \ldots \]

\[ V(I) = \pi_c \beta V(I - 1) + (1 - \pi_c) \beta V(I - 2) - \omega_B, \ldots \]

where \( B \) is the number of bottles at the end of the period.

We’ve defined the value functions as a system of equations we can solve. In particular for any inventory level \( I \) we can solve backwards until we’ve expressed \( V(I) \) in terms of \( V(1) \) and \( V(0) \). Let’s write

\[ V(I) = A_I V(1) + B_I V(0) - C_I \omega_1. \]

For \( I \leq b + 1 \). For \( I > b + 2 \) there will be terms involving \( \omega_2 \), etc, but we won’t worry about these for now. The terms \( A_I, B_I \) and \( C_I \) are positive and are functions of \( \beta \) and \( \pi_c \). We’ll focus more on these terms later but for now one useful thing to note is that

\[ A_I = \pi_c \beta A_{I-1} + (1 - \pi_c) \beta A_{I-2}, \]

and the same formula holds for \( B_I \). For \( C(I) \) the relevant equation is \( C_I = \pi_c \beta C_{I-1} + (1 - \pi_c) \beta C_{I-2} + 1. \)

Notice that if we do this we can solve for \( V(1) \) and \( V(0) \) as follows:

\[ V(0) = \frac{((1 - \pi_c)(\beta A_{N-1} - A_N) - 1)p + (C_N[(1 - \pi_c) \beta A_{N-1} - 1] - A_N(1 - \pi_c) [\beta C_{N-1} + 1]) \omega_1}{(B_N - 1)((1 - \pi_c) \beta A_{N-1} - 1) - A_N(\pi_c \beta + (1 - \pi_c) B_{N-1})} \]

Now for \( V(1) \) we get
\[ V(1) = \frac{[(1 - \pi_c)(B_n - \beta B_{N-1}) + \pi_c(1 - \beta) - 1]p + [(B_N - 1)(1 - \pi_c)(\beta C_{N-1} + 1) - C_N(\pi_c \beta + (1 - \pi_c)\beta B_{N-1})]\omega}{(B_N - 1)((1 - \pi_c)\beta A_{N-1} - 1) - A_N(\pi_c \beta + (1 - \pi_c)B_{N-1})} \]

In the steps below, we will show the value function is increasing when \( \omega_1 = 0 \). Then, we will use continuity of the value function in \( \omega_1 \) to argue it will still be increasing for small positive \( \omega_1 \) values.

Next we show that \( V(0) < 0, V(1) > V(0), \) and \( V(2) > V(1) \). Let’s start on \( V(0) \). First we will show that the numerator \( (1 - \pi_c)(\beta A_{N-1} - A_N) - 1 < 0 \). Note that we can substitute in for \( A_N \) and rewrite this as

\[ \beta(1 - \pi_c)^2(A_{N-1} - A_{N-2}) < 1. \]

It will be sufficient to show that \( A_{N-1} - A_{N-2} < 1 \). To do this I will show that \( A_I < 1 \). We can do this inductively. Recall that the equation for \( A_I \) is

\[ A_I = \beta(\pi_c A_{I-1} + (1 - \pi_c)A_{I-2}). \]

If \( A_{I-1} < 1 \) and \( A_{I-2} < 1 \) then it must be that \( A_I < 1 \). We only need to prove that \( A_2 \) and \( A_3 \) (the starting values of \( A \)) are less than 1. Note that \( A_2 = \pi_c \beta < 1 \). We can compute \( A_3 = \beta(\pi_c^2 \beta + (1 - \pi_c)) \). To show \( A_3 < 1 \) notice that if we think of \( A_3 \) as a function of \( \pi_c \) it is maximized when \( \pi_c = 1/(2\beta) \). If we plug this into \( A_3 \) then we find \( A_3 = \beta - 1/4 < 1 \). This proves the numerator of \( V(0) \) is negative. Now let’s try to prove the denominator of \( V(0) \) is positive.

We can use induction type arguments to do this too. It will be convenient to rewrite the denominator as

\[ (1 - B_N)(1 - (1 - \pi_c)\beta A_{N-1}) - A_N(\pi_c \beta + (1 - \pi_c)\beta B_{N-1}) \]

and to prove that \( 1 - B_N > \pi_c \beta + (1 - \pi_c)\beta B_{N-1} \) and \( 1 - (1 - \pi_c)\beta A_{N-1} > A_N \). Let’s start with the \( A \) inequality. This is equivalent to \( \beta A_{N-1} + \beta(1 - \pi_c)A_{N-2} < 1 \). Let’s assume that \( \beta A_{I-1} + \beta(1 - \pi_c)A_{I-2} < 1 \). We want to show this implies that \( \beta A_I + \beta(1 - \pi_c)A_{I-1} < 1 \). Note we can write

\[
\begin{align*}
\beta A_I + \beta(1 - \pi_c)A_{I-1} &= \beta^2 \pi_c A_{I-1} + \beta^2(1 - \pi_c)A_{I-2} + \beta(1 - \pi_c)A_{I-1} \\
&= \beta(1 + \pi_c(\beta - 1))A_{I-1} + \beta^2(1 - \pi_c)A_{N-2} \\
&< \beta A_I + \beta(1 - \pi_c)A_{I-1} < 1
\end{align*}
\]
Then we just have to prove the base case which is that $\beta^2(\pi_e^2\beta + (1 - \pi_c)) + \beta^2(1 - \pi_c)\pi_c < 1$. This is true since we can reduce the inequality to $\beta^2(1 + \pi_e^2(\beta - 1)) < 1$.

Now let’s do the $B$ inequality. We will rewrite the inequality as $\pi_c\beta + (1 - \pi_c)\beta B_{N-1} + B_N < 1$. Assume that $\pi_c\beta + (1 - \pi_c)\beta B_{I-2} + B_{I-1} < 1$. Then we can write

$$B_I + \beta(1 - \pi_c)\beta B_{I-1} + \pi_c\beta = \beta\pi_c B_{I-1} + \beta(1 - \pi_c)B_{I-2} + (1 - \pi_c)\beta B_{I-1} + \beta\pi_c$$

$$= \beta B_{I-1} + \beta(1 - \pi_c)B_{I-2} + \beta\pi_c$$

$$< \pi_c\beta + (1 - \pi_c)\beta B_{I-2} + B_{I-1} < 1$$

Again we have to prove the initial case, which is $\pi_c\beta + (1 - \pi_c)^2\beta^2 + \pi_c(1 - \pi_c)\beta^2 < 1$. We can reduce this inequality to $\beta(\pi_c + \beta(1 - \pi_c)) < 1$, and it is easy to see this will be true.

Next we want to show $V(1) > V(0)$. Since we have proven that the denominators of these terms are positive we need to show that

$$(1 - \pi_c)(B_N - \beta B_{N-1}) + \pi_c(1 - \beta) - 1 > (1 - \pi_c)(\beta A_{N-1} - A_N) - 1$$

We can rewrite this inequality as

$$(1 - \pi_c)(\beta(A_{N-1} + B_{N-1}) - (A_N + B_N)) < \pi_c(1 - \beta)$$

It is sufficient to show that $\beta(A_{N-1} + B_{N-1}) - (A_N + B_N) < 0$. Inductive arguments can be used here. First consider the base cases. The base case at 2 boils down to showing $\beta^2 < (1 - \pi_c)\beta + \pi_c\beta^2$.

This inequality is equivalent to $\beta < (1 - \pi_c) + \pi_c\beta$ or $\beta < 1$ (we’d actually get equality is $\pi_c = 1$ but I don’t think that is a problem).

Next, we also need to show that $\beta(A_3 + B_3) < A_4 + B_4$. This is straightforward. We know $A_4 + B_4 = \pi_e^2\beta^2 + (1 - \pi_e^2)\beta^2$. We need to show that $(1 - \pi_c)\beta^2 + \pi_c\beta^3 < \pi_e^2\beta^2 + (1 - \pi_e^2)\beta^2$. This inequality reduces to $(\beta - 1)\pi_c < (\beta - 1)\pi_e^2$. It is easy to see this is true since $\pi_c$ is a fraction and $\beta < 1$.

To finish we use complete induction. Suppose for all $n < I$ it is the case that $\beta(A_{n-1} + B_{n-1}) < A_n + B_n$. Then we can show

$$\beta(A_{I-1} + B_{I-1}) < A_I + B_I$$

$$\iff \beta(\pi_c(A_{I-2} + B_{I-2}) + (1 - \pi_c)(A_{I-3} + B_{I-3})) < \beta(\pi_c(A_{I-1} + B_{I-1}) + (1 - \pi_c)(A_{I-2} + B_{I-2}))$$

$$\iff \beta(\pi_c(A_{I-2} + B_{I-2}) + (1 - \pi_c)(A_{I-3} + B_{I-3})) < \pi_c(A_{I-1} + B_{I-1}) + (1 - \pi_c)(A_{I-2} + B_{I-2})$$
Our induction assumption implies \( \beta(\pi_c(A_{I-2} + B_{I-2}) < \pi_c(A_{I-1} + B_{I-1}) \) and \( \beta((1 - \pi_c)(A_{I-3} + B_{I-3}) < (1 - \pi_c)(A_{I-2} + B_{I-2}) \).

We also need to show \( V(2) > V(1) \). Here is how to do it. We need to prove that \( \beta\pi_c V(1) + \beta(1 - \pi_c) V(0) > V(1) \). Using our formulas above and the fact that the denominator of the \( V \)'s is negative we can reduce the inequality to

\[
\beta^2(1 - \pi - c)^2 A_{N-1} + (1 - \beta\pi_c)(1 - \pi_c)\beta B_{N-1} > \beta(1 - \pi_c)^2 A_N + (1 - \beta\pi_c)(1 - \pi_c) B_N + (1 - \beta\pi_c)(\pi_c(1 - \beta) - 1) + \beta(1 - \pi_c) \\
\iff (1 - \pi_c)([\beta(1 - \pi_c)(\beta A_{N-1} - A_N)] + (1 - \beta\pi_c)[\beta B_{N-1} - B_N]) > (\beta - 1)(\pi_c^2\beta - \pi_c + 1)
\]

The right side of the inequality is negative so to make things a bit easier I’m going to multiply both sides by -1 and work with an upper bound. Suppose that

\[
(1 - \pi_c)([\beta(1 - \pi_c)(A_I - \beta A_{I-1})] + (1 - \beta\pi_c)[B_I - \beta B_{I-1}]) < (1 - \beta)(\pi_c^2\beta - \pi_c + 1)
\]

for all \( I \leq N \). Then we can write the \( N + 1 \) case as

\[
(1 - \pi_c)([\beta(1 - \pi_c)(A_{N+1} - \beta A_N)] + (1 - \beta\pi_c)[B_{N+1} - \beta B_N]) \\
= (1 - \pi_c)([\beta(1 - \pi_c)(\beta \pi_c A_N + \beta(1 - \pi_c) A_{N-1} - \beta(\beta \pi_c A_{N-1} + \beta(1 - \pi_c) A_{N-2})] + \\
(1 - \beta\pi_c)(\beta(1 - \pi_c)(\beta \pi_c B_N + \beta(1 - \pi_c) B_{N-1} - \beta(\beta \pi_c B_{N-1} + \beta(1 - \pi_c) B_{N-2})) \\
= \beta \pi_c([\beta(1 - \pi_c)(A_N - \beta A_{N-1}) + (1 - \beta\pi_c)(B_N - \beta B_{N-1})] \\
\beta(1 - \pi_c)(1 - \beta)(\pi_c^2\beta - \pi_c + 1) + \beta(1 - \pi_c)(1 - \beta)(\pi_c^2\beta - \pi_c + 1) \\
< (1 - \beta)(\pi_c^2\beta - \pi_c + 1)
\]

Last we need to show the base case, for 2, 3 and 4. Recall that \( A_2 = \pi_c\beta, B_2 = (1 - \pi_c)\beta, A_3 = \pi_c^2\beta^2 + (1 - \pi_c)\beta, B_3 = \pi_c(1 - \pi_c)\beta^2, A_4 = \pi_c^3\beta^3 + 2\pi_c(1 - \pi_c)\beta^2, \) and \( B_4 = \pi_c^2(1 - \pi_c)\beta^3 + (1 - \pi_c)^2\beta^2 \).

First we prove the case for time periods 2 and 3. It turns out that the left side of the inequality is

\[
(1 - \pi_c)[\beta(1 - \pi_c)(\pi_c^2\beta^2 - \pi_c\beta^2 + (1 - \pi_c)\beta) + (1 - \beta\pi_c)(1 - \pi_c)\beta^2(\pi_c - 1)] = 0
\]
So that part follows. Then we do the same for periods 3 and 4. To start notice that we can write the left hand side of the inequality as

\[ \beta^2(1 - \pi_c)^2[\beta(1 - \pi_c)(2\pi_c - 1 - (1 - \pi_c)\beta) + (1 - \beta\pi_c)(\pi_c^2\beta + 1 - \pi_c - \pi_c\beta)] \]

We want to simplify the stuff inside the square brackets. If you do a bunch of algebra you can reduce this to

\[
\pi_c\beta - \beta + 2\pi_c\beta^2 - \beta^2 + 1 - \pi_c - \pi_c^3\beta^2
= -(1 - \pi_c)\beta + (-\pi_c^3 + 2\pi_c - 1)\beta^2 + (1 - \pi_c)
\]

We can factor the term multiplying \( \beta^2 \) into \( 1 - \pi_c \) and \( \pi_c^2 + \pi_c - 1 \). So then the inequality becomes

\[
\beta^2(1 - \pi_c)^3(1 - \beta + (\pi_c^2 + \pi_c - 1)\beta^2) < (1 - \beta)(\pi_c^2\beta - \pi_c + 1)
\]

Note that we can show that \( \pi_c^2 + \pi_c - 1 < 0 \). This is a quadratic that is -1 when \( \pi_c \) is 0, is 0 when \( \pi_c \) is 1, and is strictly increasing in that interval. So it is sufficient to show that \( \beta^2(1 - \pi_c)^3 < \pi_c^2\beta - \pi_c + 1 \). This is straightforward because \( \beta^2(1 - \pi_c)^3 < 1 - \pi_c < \pi_c^2\beta - \pi_c + 1 \). The right inequality holds since \( \pi_c^2\beta > 0 \) and the left one holds since \( \beta(1 - \pi_c)^2 < 1 \).

To show the value function is increasing in inventory for all \( I \) we can use the fact we showed \( V(1) > V(0) \) and apply complete induction. We assume that our statement is true for all inventory levels \( n \) less than \( I \). In other words, if \( n < I \) we assume that \( V(n) > V(n - 1) \) (in particular it means that \( V(I - 1) > V(I - 2) > V(I - 3) > ... \)). We want to show that this implies that \( V(I) > V(I - 1) \).

Let’s start by noticing that \( V(I - 1) > V(I - 2) \) implies the following:

\[ V(I - 1) - V(I - 2) = (A_{I-1} - A_{I-2})V(1) + (B_{I-1} - B_{I-2})V(0) > 0 \]

There are two things we will need to complete the proof. First, notice that

\[
A_I - A_{I-1} = \beta\pi_cA_{I-1} + \beta(1 - \pi_c)A_{I-2} - A_{I-1}
= \beta\pi_c(A_{I-1} - A_{I-2}) + \beta(A_{I-2} - A_{I-1})
\]
and the same holds true for the $B$ series. The second thing we want to show is that $\beta V(I - 2) \geq V(I - 1)$. Note that

$$V(I - 1) = \beta \pi c V(I - 2) + \beta (1 - \pi c) V(I - 3)$$

So we want to argue that

$$V(I - 2) \geq \pi c V(I - 2) + (1 - \pi c) V(I - 3).$$

We assumed that $V(I - 2) > V(I - 3)$, so $\pi c V(I - 2) + (1 - \pi c) V(I - 3)$ is maximized when $\pi c = 1$. Then equality holds. Otherwise the inequality must be strict since we are increasing weight on $V(I - 3)$.

Now notice that we can write $V(I) - V(I - 1)$ as

$$V(I) - V(I - 1) = (A_I - A_{I-1})V(1) + (B_I - B_{I-1})V(0)$$

$$= (\beta \pi c(A_{I-1} - A_{I-2}) + \beta(A_{I-2} - A_{I-1})) V(1) + (\beta \pi c(B_{I-1} - B_{I-2}) + \beta(B_{I-2} - B_{I-1})) V(0)$$

$$= \beta \pi c(V(I - 1) - V(I - 2)) + \beta V(I - 2) - V(I - 1)$$

Our induction assumption was $V(I - 1) > V(I - 2)$ so $\beta \pi c(V(I - 1) - V(I - 2)) > 0$. Additionally we proved above that $\beta V(I - 2) - V(I - 1) > 0$. Hence, $V(I) - V(I - 1) > 0$ and the lemma is proved.

Last we note that all the value functions are continuous in $\omega_1$. This fact will imply that the set of inequalities we proved above will still hold for small values of $\omega_1$.

### 11.2 Proof of Lemma 2

We now prove that the proposed policy is optimal. To do this we will show that the value functions we derived in the last lemma are consistent with the optimality conditions. This implies policies derived from them are optimal.

We can prove this lemma inductively. We first note that the consumer will always make a purchase if she is going to run out. To see this, note that the value of purchasing will be $-p - \omega_1 + \beta V(b - c_{it})$, while the value from running out and not purchasing is $-\nu + \beta V(0)$. Since the value function is increasing in inventory due to Lemma 1, and we have assumed $p < \nu - \omega_1$, the payoff from purchasing is higher than the payoff from running out. Before we continue we note that this part of the proof is the only place where we use Lemma 1. Lemma 1 is a bit stronger than we need
we really only need it to be the case that \( V(b - 2) - V(0) \geq 0 \) and \( V(b - 1) - V(0) \geq 0 \). These two conditions are harder to prove, but we think they should be true under weaker conditions than what is required to prove Lemma 1.

Next we want to show that the payoff from not purchasing is higher than the payoff from purchasing when the consumer will not run out. It is sufficient to demonstrate that

\[
\beta V(0) > -p - \omega_1 + \beta V(b),
\]

for \( I = 0 \) and

\[
\beta V(I) - \omega_1 > -p - \omega_2 + \beta V(b + I),
\]

for \( I \geq 0 \). We begin by showing that \( \beta V(0) > -p + \beta V(b) \) and \( \beta V(1) - \omega_1 > -p - \omega_2 + \beta V(b + 1) \), and then use induction after. To show the first inequality note that following the optimal policy at 0 inventory it must be the case that

\[
V(0) = -p + \beta \pi_c V(b - 1) + \beta(1 - \pi_c) V(b - 2) - \omega_1 = -p + V(b).
\]

Thus \( V(b) - V(0) = p \). Since \( \beta < 1 \) and \( p + \omega_1 \) is positive, it must be that \( \beta(V(b) - V(0)) < p \). It is easy to see this implies \( \beta V(0) > -p - \omega_1 + \beta V(b) \).

Similar logic can be used to show \( \beta V(1) - \omega_1 > -p - \omega_2 + \beta V(b + 1) \). First note that we can write \( V(1) \) as follows:

\[
V(1) = \beta \pi_c V(0) + \beta(1 - \pi_c) V(b - 1) - (1 - \pi_c)p - (1 - \pi_c) \omega_1
\]

\[
= \beta \pi_c V(b - p) + \beta(1 - \pi_c) V(b - 1) - (1 - \pi_c)p - (1 - \pi_c) \omega_1
\]

\[
= V(b + 1) - (1 - \pi_c(1 - \beta))p + \pi_c \omega_1
\]

The inequality is true since \( \beta(V(b + 1) - V(1)) < (1 - \pi_c(1 - \beta))p - \pi_c \omega_1 < p + \omega_2 - \omega_1 \).

Now we will use induction to prove optimality generally. Suppose that the inequality is true up to \( b + I \), in other words that \( \beta(V(b + n) - V(n)) < p + \omega_B - \omega_{B-1} \) for \( 0 \leq n \leq I \). We will work with the value function difference \( \beta(V(b + n + 1) - V(n + 1)) \). First, assume no storage cost change between \( I + 1, I \), and \( I - 1 \). Then we can write this difference as

\[
\beta(V(b + I + 1) - V(I + 1)) = \beta \pi_c \beta(V(b + I) - V(I)) + (1 - \pi_c) \beta(V(b + I - 1) - V(I - 1)) - (\omega_B - \omega_{B-1})
\]

\[
< \beta p
\]

\[
< p + \omega_B - \omega_{B-1}.
\]
where the second inequality follows from the induction assumption. Note that our assumption that storage costs are weakly increasing is important here.

Suppose there is a storage cost change between $I + 1$ and $I$. Then the inequalities become

\[
\beta(V(b + I + 1) - V(I + 1)) = \beta((\pi_c \beta(V(b + I) - V(I)) - \pi_c (\omega_B - \omega_{B-1}) + (1 - \pi_c)(V(b + I - 1) - V(I - 1)) - (1 - \pi_c)(\omega_{B-1} - \omega_{B-2}))
\]

\[
< \beta(p + (1 - \pi_c)(\omega_B - \omega_{B-1}) + \pi_c(\omega_{B-1} - \omega_{B-2}))
\]

\[
< p + \omega_B - \omega_{B-1},
\]

The last inequality will follow as a result of weak convexity of the storage cost function. The case where storage costs change between $I$ and $I - 1$ is similar. Thus, the policy proposed is optimal. This concludes the proof of Lemma 2.

12 Appendix: Artificial Data Experiment with Continuous Inventory

In this section we describe an additional artificial data experiment where inventory is continuous rather than discrete. We solve for consumer value functions and simulate choices in a market where the utility parameterization is based on the values in Table ???. We simulate a dataset of 1000 consumers, for 700 periods. We assume in the initial simulation period that all consumers start with inventory of 0. In a data set tracking the behavior of real consumers, consumers will have been making purchases prior to the beginning of the data collection, so initial inventories will be unknown. To mimic this in the simulated data we remove the first 200 periods and estimate the model parameters using the final 500 periods. We estimate the model using the nested fixed point algorithm (Rust 1987) on the simulated data to see how well we can recover the model parameters. Note that when we estimate the model, we need to construct initial inventories in order to evaluate the likelihood. To do this we take the observed 500 periods and split them in half. We assume that in the initial period all consumers have zero inventory, simulate consumption rates, and compute inventory in period 250 as the sum of observed purchases minus consumption rates (where we enforce the restriction that inventory is greater than or equal to zero at each period). We evaluate the likelihood on the final 250 periods for each consumer. To verify that 250 periods is enough to accurately simulate initial inventories, we simulate the model and compute moments of the inventory distribution over time, finding that the inventory distribution looks like
it becomes stationary after about 50 periods. We use a simplex algorithm, with a penalty function to enforce nonnegativity constraints, to maximize the likelihood.\footnote{We experimented with derivative based methods as well but found the simplex reached slightly higher likelihood values. Occasionally the derivative based methods would stop at the starting points.}

We run the artificial data experiment for 8 different parameterizations of the model. The results of the experiment for the first 6 parameterizations are shown in Table 8. The upper half of the table and first 3 columns of the table show, respectively, the estimated parameters, standard errors, and true parameters for the basic specification from the last section. The results here suggest that the estimation works well - the estimated parameter values are close to the true values. The next 2 sets of 3 columns, under the headings $\beta = 0.9$ and $\beta = 0.99$, show the results for these different $\beta$ values. Again, the estimation works well. Turning to the lower half of the table we see under the heading $\beta = 0.6$ that the estimation still works well for this value of $\beta$. The story changes somewhat for the last two sets of estimates, under $\beta = 0.3$ and $\alpha = 0.02$. When $\beta = 0.3$, the estimates of $\beta$, the stockout cost $\nu$, and the storage cost $\omega$ become significantly worse. As we discussed in the last section, consumers behave in a manner that is very close to myopic for low values of $\beta$, and so identification of the discount factor becomes more difficult. In general the storage cost and stockout costs will get harder to identify for low values of $\beta$. To see why, consider how a myopic consumer will behave in a stockpiling model like ours. Such a consumer will only purchase when she runs out of the product, and will be very unlikely to purchase more than a single unit (in a model with no logit error she would never purchase more than 1 unit). The stockout cost will only affect consumer choices when a consumer has no inventory left at all. The stockout cost will be identified by price variation that occurs when a consumer runs out. If the disutility from paying is higher than the stockout cost the consumer will not purchase when she runs out; if the price is sufficiently low she will be induced to purchase. In contrast, for a forward-looking consumer the stockout cost will affect purchase decisions prior to running out, since consumers anticipate the stockout cost; thus the type of price variation necessary to identify the stockout cost (the tendency to purchase in response to low prices) will be important over more periods than just ones where a consumer runs out. This logic also implies that if consumers are relatively price insensitive, the stockout cost could become hard to identify. Indeed, this is exactly what we observe when we try to recover the model parameters for a value of $\alpha = 0.02$: the estimated stockout cost is quite different from the true value. Fortunately the discount factor is still well-identified in this situation: this is so because the discount factor is primarily identified by the change in the choice probability over time, rather than by consumer response to price variation.

Table 9 shows simulated results for a quadratic storage cost, rather than the piecewise linear cost.
Table 8: Results of Artificial Data Experiment (Quasilinear Storage Cost)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta = 0.95$</td>
<td></td>
<td></td>
<td>$\beta = 0.99$</td>
<td></td>
<td></td>
<td>$\beta = 0.9$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.001</td>
<td>(0.002)</td>
<td>0</td>
<td>0.001</td>
<td>(0.002)</td>
<td>0</td>
<td>0.001</td>
<td>(0.003)</td>
<td>0</td>
</tr>
<tr>
<td>$\tau$</td>
<td>2.012</td>
<td>(0.001)</td>
<td>2</td>
<td>2.011</td>
<td>(0.001)</td>
<td>2</td>
<td>2.014</td>
<td>(0.002)</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.05</td>
<td>(4.1e-04)</td>
<td>0.05</td>
<td>0.05</td>
<td>(4.4e-04)</td>
<td>0.05</td>
<td>0.05</td>
<td>(4.0e-04)</td>
<td>0.05</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.983</td>
<td>(0.023)</td>
<td>1</td>
<td>0.964</td>
<td>(0.029)</td>
<td>1</td>
<td>0.965</td>
<td>(0.019)</td>
<td>1</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.946</td>
<td>(0.002)</td>
<td>0.95</td>
<td>0.985</td>
<td>(0.002)</td>
<td>0.99</td>
<td>0.897</td>
<td>(0.003)</td>
<td>0.9</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.099</td>
<td>(0.003)</td>
<td>0.1</td>
<td>0.101</td>
<td>(0.002)</td>
<td>0.1</td>
<td>0.097</td>
<td>(0.004)</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>$\beta=0.6$</td>
<td></td>
<td></td>
<td>$\beta=0.3$</td>
<td></td>
<td></td>
<td>$\alpha=0.02$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.004</td>
<td>(0.007)</td>
<td>0</td>
<td>3.9e-04</td>
<td>(0.006)</td>
<td>0</td>
<td>0.001</td>
<td>(0.002)</td>
<td>0</td>
</tr>
<tr>
<td>$\tau$</td>
<td>2.004</td>
<td>(0.005)</td>
<td>2</td>
<td>2.002</td>
<td>(0.004)</td>
<td>2</td>
<td>2</td>
<td>(0.001)</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.05</td>
<td>(3.4e-04)</td>
<td>0.05</td>
<td>0.05</td>
<td>(3.9e-04)</td>
<td>0.05</td>
<td>0.02</td>
<td>(1.4e-04)</td>
<td>0.02</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.967</td>
<td>(0.045)</td>
<td>1</td>
<td>0.876</td>
<td>(0.088)</td>
<td>1</td>
<td>0.565</td>
<td>(0.045)</td>
<td>1</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.615</td>
<td>(0.023)</td>
<td>0.6</td>
<td>0.407</td>
<td>(0.064)</td>
<td>0.3</td>
<td>0.944</td>
<td>(0.008)</td>
<td>0.95</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.086</td>
<td>(0.009)</td>
<td>0.1</td>
<td>0.071</td>
<td>(0.013)</td>
<td>0.1</td>
<td>0.11</td>
<td>(0.011)</td>
<td>0.1</td>
</tr>
</tbody>
</table>

function we are currently using. We assume that the linear part of storage costs is 0, and estimate the quadratic coefficient $\omega$. Column 1 shows the results with the exclusion restriction. In Column 2, we estimate the parameters in a situation where storage costs are assumed to be continuous rather than discrete (we simulate the model under continuous storage costs and maintain that assumption when we estimate the model). All the parameters are still well-identified, suggesting that the nonlinear restrictions from the dynamic model provide some identification. However, it is notable that the standard errors are higher (and sometimes very much higher as in the case of the consumption rates). This suggests that the exclusion restrictions provide identifying variation beyond the functional form restrictions of the model.
### Table 9: Estimation with and without the Exclusion Restriction (Quadratic Storage Cost)

<table>
<thead>
<tr>
<th>Param</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Truth</th>
<th>Estimate</th>
<th>S.E.</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Discrete</strong></td>
<td></td>
<td></td>
<td><strong>Continuous</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$c$</td>
<td>0.001</td>
<td>(0.006)</td>
<td>0</td>
<td>0.001</td>
<td>(0.024)</td>
<td>0</td>
</tr>
<tr>
<td>$\bar{c}$</td>
<td>2.013</td>
<td>(0.006)</td>
<td>2</td>
<td>2.013</td>
<td>(0.024)</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.05</td>
<td>(4.1e-04)</td>
<td>0.05</td>
<td>0.05</td>
<td>(4.3e-04)</td>
<td>0.05</td>
</tr>
<tr>
<td>$\nu$</td>
<td>0.975</td>
<td>(0.025)</td>
<td>1</td>
<td>0.987</td>
<td>(0.028)</td>
<td>1</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.946</td>
<td>(0.002)</td>
<td>0.95</td>
<td>0.945</td>
<td>(0.002)</td>
<td>0.95</td>
</tr>
<tr>
<td>$\omega$</td>
<td>0.005</td>
<td>(1.8e-04)</td>
<td>0.005</td>
<td>0.005</td>
<td>(2.0e-04)</td>
<td>0.005</td>
</tr>
</tbody>
</table>