Job Market Paper

High-dimensional inventories and consumer dynamics: demand estimation for fast moving consumer goods

Alan Crawford

November, 2017

[Please click here for the most recent version]

This paper develops a high-dimensional dynamic discrete-continuous demand model for storable fast moving consumer goods. Assumptions of existing models are relaxed while retaining computational tractability. As a result, the model captures rich inter- and intra-temporal substitution patterns, allows for a detailed understanding of dynamic consumer behaviour, and provides a framework with wide applicability. To estimate and solve the dynamic demand model, I use techniques from approximate dynamic programming, large-scale dynamic programming in economics, machine learning, and statistical computing. In this paper I apply the model to the UK laundry detergent sector using household level purchase data.

Acknowledgements: I wish to thank the Economic Social Research Council for funding this research and Kantar Worldpanel for the use of their data. I would also like to thank my supervisor, Lars Nesheim, for his advice and guidance. I would also like to thank Dennis Kristensen and Aureo de Paula for helpful comments and suggestions. Data supplied by Kantar Worldpanel. The use of Kantar Worldpanel data in this work does not imply the endorsement of Kantar Worldpanel in relation to the interpretation or analysis of the data. All errors and omissions remain the responsibility of the author.

1 Introduction

Storable fast moving consumer goods (FMCGs) account for a large portion of household grocery expenditures. In the UK, the storable FMCG market is worth in excess of £1 trillion. These include alcohol, tinned produce, frozen produce, table sauces, condiments, personal hygiene products, household cleaning products, and pet food.

There are two key features of storable goods. First, they are typically sold at a constant price for sustained periods of time interspersed with frequent, stochastic, temporary sales.[1] While the timing of promotions is uncertain, prices exhibit persistence. Through repeated shopping trips, households form expectations over future prices and factor them into their purchase decisions.

Second, the durability of storable FMCGs enables them to be held in inventory to be consumed at a later date. As a result, households make infrequent purchases to meet both current and future consumption needs. The process of building up stocks with purchases and depleting them through consumption creates inter-temporal links in demand behaviour.

In light of these considerations, inventories are a key determinant of demand dynamics. When inventories are high, households can service current consumption with existing inventory that occupies limited storage space. The combination of these factors results in low demand for new purchases. At the other extreme, when a household's stocks are depleted, demand for new stock is high because they need to make purchases to consume. At intermediate levels of inventory, households can choose to make purchases for both consumption and storage. In particular, they may accelerate purchases in response to price promotions.

Problematically for demand estimation, inventories are unobserved by researchers.[2] Because current prices and inventories are a function of past prices, the omission of inventories leads to price endogeneity. In turn, elasticities and welfare analysis based on a static demand estimation are biased. This problem could be resolved if instruments that are correlated with current prices but uncorrelated with past prices were available.[3] However, since observed prices are serially correlated, finding such instruments is challenging - if not impossible.

An alternative is to integrate out over inventories during estimation - a computationally intensive procedure requiring simulation of sequences of purchases, consumption and inventories. This approach is further complicated by the fact that households have the option to purchase, store and consume many different variants of the good. For example, in the UK laundry detergent industry, households' choice set contains approximately 100 purchase options with over 35 different types of detergent

[1] Erdem et al. (2003); Hendel and Nevo (2006a,b); Nevo and Hendel (2012); Osborne (2013); Wang (2012, 2013) all report that prices typically exhibit these features.

[2] Even though detailed household level purchase diary data records price, quantity purchased, product characteristics and date of all purchases for each household in the survey, inventories cannot be constructed without data on initial inventories and consumption.

[3] With these instruments, marginal utility of income could be consistently estimated using a micro-BLP procedure.

in any given week.

Even without the unobserved inventories, the curse of dimensionality is severe when solving and estimating a high-dimensional dynamic choice model. The need to integrate out over the high dimensional inventory state space further exacerbates this sizeable computational challenge.

Faced with these issues, existing dynamic demand models for storable FMCGs have sought to limit the model's flexibility to reduce computational resources needed to estimate it. In general, there have been two approaches.

One approach imposes restrictions on the model so that the consumer's decision can be split into a static brand choice and a dynamic quantity choice (Hendel and Nevo (2006a); Wang (2015); Osborne (2017)). Building on an approach pioneered by Melnikov (2000, 2013), the brand-size split is further leveraged by assuming that consumers' price expectations are captured by the evolution of the ex-ante expected utility of purchasing particular sizes.

The other approach allows for both the brand and quantity choice to be dynamic. The price process varies over brands or sizes and is consistent with observed price movements (Erdem et al. (2003); Sun (2005)).

Both approaches adopt restrictions on the functional form of utility from consumption. Further, if the number of choices entering the choice set is large, neither of these approaches yield computationally tractable demand models. While successful, the range of applications is highly limited. Indeed, in many industries, products are available in many different varieties and sizes and the choice sets are high-dimensional.

UK laundry detergent is one such industry and is the subject of this paper. It is chosen because it highlights many of the challenges that arise when estimating dynamic demand models for storable goods. First, as noted above, the choice set is high-dimensional - there are around 100 products in the choice set in each week.[4] Second, differentiation of detergents affects their quality, ease of use, and costs of storage. Third, it exhibits the promotional price patterns that are often observed in many storable good industries.

This paper presents two key innovations to alleviate the curse of dimensionality. First, faced with the cognitively infeasible task of forecasting a high dimensional time series of prices, boundedly rational households are assumed to use a low-rank statistical approximation to model price dynamics. Through repeated shopping visits, households are assumed to develop a hedonic model to reflect the cross-sectional distribution of prices and a low dimensional factor model to capture price dynamics.[5]

Second, when deciding which, if any, detergent to purchase households are assumed to use an approximate solution to their high-dimensional dynamic choice problem. This too can be viewed as a boundedly-rational approach to making cognitively challenging decisions.

Specifically, households are assumed to make use of the fact that they stock no more than a few products at a time. In line with this observation, when making purchase

[4] There are eight major brands, available in five different formats sold in many different pack sizes (i.e. over 13 pack sizes are needed to cover 95 percent of sales).

[5] Bai and Ng (2008) provide an overview of factor models.

decisions households are assumed to only consider the value of the stock they can consume. Therefore a good decision rule needs only to keep track of a handful of detergent inventories at any one time and can use a low-dimensional representation of the large inventory state space with minimal loss of information.

Beneficially, these dimension reduction strategies impose fewer restrictions than existing models. As such, this modelling framework can build on the seminal work of Erdem et al. (2003) and Hendel and Nevo (2006a).

First, the price forecasting model developed in this paper builds on aspects of the different approaches taken by Hendel and Nevo (2006a) and Erdem et al. (2003).

As noted above, Hendel and Nevo (2006a) use the evolution of ex-ante expected utility of consuming each size to model price dynamics. Using this approach all brands of the same size are assumed to have identical price processes. In this case, inter-temporal substitution patterns produced by their model reflect consumer responses to expected changes to quality-adjusted price indices for each size.

Beneficially, consumers can take into account differences in price dynamics related to the size of products. Indeed, in the UK laundry detergent industry smaller products tend to be promoted more frequently, than larger products. However, one drawback of this approach is that unless the quality-adjusted price indices are dominated by specific brands, brand related price dynamics do not influence consumer purchases or consumption.

In reality, it is likely that through repeated shopping visits, consumers understand promotional pricing patterns vary over brands. While some brands are never promoted, others are promoted relatively frequently. They also observe that promotions of competing products tend to be sequenced as a result of what they perceive to be manufacturers' responses to rivals' promotions.

To build on our understanding of how such price dynamics affect consumer demand, it is therefore desirable to enable consumers to respond to brand-specific price dynamics observed in the data. Indeed, this appears to be one of the motivating factors behind the design of the price process used in Erdem et al. (2003).

In their model of the US ketchup market, they estimate brand specific price processes. However, to limit the dimension of each price process, they assume that the price movements of the brand's most popular pack size are mirrored by the other sizes. However, while this restriction may be suitable in the US ketchup industry - it may not be appropriate in other FMCG industries.

In my forecasting model, consumer expectations can differ over both the size and brand of the product. This is possible because my statistical price forecasting model uses a low rank approximation to product-specific asynchronous promotional price patterns observed in the data.

Second, I extend existing models by allowing for high-dimensional choice sets. As noted above, existing models require that the number of alternatives in the dynamic discrete choice problem is low-dimensional. This is because the number of state space variables needed to forecast prices is directly linked to the size of the choice set.

For Hendel and Nevo (2006a) this issue arises because the state space includes a quality-adjusted price index for each pack size. Therefore computational benefits of imposing utility restrictions are quickly eroded by the curse of dimensionality when there are more than a few different sizes (i.e. more than 4 or 5). In Erdem et al. (2003) the dimension of the state space increases with the number of brands. These constraints have limited the applicability of these existing models to many storable FMCG industries.

However, my forecasting model does not directly depend on the number of brands or pack sizes in the choice set to reduce the dimension of the price forecasting problem. Instead, it does so by exploiting underlying correlations in observed price series. As a result, the link between the cardinality of the choice set and the dimension of the state space is broken. Without these constraints, I can incorporate high-dimensional choice sets with many brands and sizes. Moreover, I do so without adding more restrictions to the utility function.[6]

Finally, extending the approaches of Erdem et al. (2003) and Hendel and Nevo (2006a), I allow consumption to be endogenous and multi-dimensional.[7] When combined with additional flexibility of my price forecasting model, this approach to modelling consumption enables households to respond to expected price changes by altering consumption and/or purchases.

Taken together, these advances provide a dynamic demand modelling framework that can be used to develop increasingly realistic models of consumer behaviour. In turn, delivering new insights into the nature of competitive interactions between firms in these industries and aiding antitrust policy analysis.

To estimate and approximate the solution to the dynamic demand model, I use techniques from approximate dynamic programming (ADP). ADP combines tools from statistics and machine learning to approximate the solution to computationally intractable dynamic programming problems. It encompasses a wide variety of techniques from a collection of disparate fields that have developed specific approaches to approximate solutions to complex dynamic programs they encounter.[8]

There exists a nascent literature where ADP methods have been used to estimate dynamic models in economics. Hendel and Nevo (2006a), Sweeting (2013) and Fowlie et al. (2016) use parametric policy function iteration described by Benitez-Silva et al.

[6]In contemporary work, Osborne (2017) builds on Hendel and Nevo (2006a) to allow for multiple product sizes. To achieve this he adds more restrictions to the utility function. By ruling out quantity discounts and assuming that flow payoffs can be rescaled by the quantity purchased of a given brand, Osborne (2017) reduces the state vector to a single expected utility state. As a result all brands and sizes are assumed to follow the same price process. The benefit of this approach is that he makes the expected utility of consuming a given brand continuous in the amount purchased. Therefore, these additional restrictions can be used to circumvent computational costs associated with Hendel and Nevo (2006a) inclusive value approach when there are many sizes.

[7]In Erdem et al. (2003), consumption is multi-dimensional and assumed to be exogenous. To circumvent the problems of tracking a high dimensional inventory, Erdem et al. (2003) assume that consumption is proportionately drawn from all stocks. Combined with a linear utility specification, this reduces the inventory state space to just two dimensions. In Hendel and Nevo (2006a) consumption is endogenous and utility from consumption is assumed to be the same for all products. Therefore only aggregate inventories are needed in the state space.

[8]Indeed, some of the techniques used are closely linked to generalised stochastic search algorithms that have been used to solve large-scale dynamic programs in economics (see Maliar and Maliar (2014); Judd et al. (2011)).

(2000) - an early ADP algorithm. Arcidiacono et al. (2012) show how to use sieve value function iteration to estimate and approximate the solution to dynamic single agent models with large-state spaces - an approach closely related to the ADP methods used in this paper. Other ADP techniques have also been used to approximate solutions to large scale dynamic games (see Farias et al. (2012)).

One of the key concepts of ADP is that dynamic models should reflect the reality of the decision environment as closely as possible. Then, within this detailed dynamic model, agents are assumed to make near-optimal decisions using an approximate solution to their choice problem. This mirrors the boundedly rational approach households are assumed to use to solve their high-dimensional dynamic choice problem.

While the dimension of the price and inventory state space is reduced, the state space is still moderately sized. To help mitigate the curse of dimensionality associated with moderately sized dynamic programs the value function is approximated using a Smolyak polynomial (see Judd et al. (2014)). Beneficially, the basis functions of this family of polynomials grow polynomially - not exponentially - in the size of the state space.

The approximation to the solution of the household's choice problem is characterised by the coefficients of this polynomial. However, if the optimal solution does not lie in the space spanned by the approximating polynomial, then the fixed point of the Bellman equation evaluated using the approximate solution may not exist.

An alternative approach is to solve a modified version of the Bellman equation in which the optimal solution is projected onto the space spanned by the basis functions of the Smolyak polynomial. In this case, the approximate solution to the household's dynamic choice problem is the set of coefficients that define the fixed point of a projected version of the Bellman equation.

As highlighted by Bertsekas (2011b), the application of standard dynamic programming techniques will not necessarily deliver the fixed point of a projected Bellman equation.[9] Therefore, to find this fixed point, I use an ADP algorithm called λ-policy iteration (Bertsekas (2015)) designed to solve projected Bellman equations. Like exact policy iteration, this algorithm repeatedly applies two steps: (i) policy evaluation, and (ii) policy improvement.

The value of using the policy being evaluated in step (i) is calculated using simulation. To find the coefficients that represent the value of following the policy, the sum of squares of the Bellman equation residuals visited along the simulation trajectory are minimised using stochastic projected gradient methods (Bertsekas (1999); Parikh et al. (2014)).[10] The policy improvement step uses the envelope condition method (Maliar and Maliar (2013)). The solution to the projected Bellman equation results from iterative application of these two steps until convergence or until a pre-specified large number of iterations are completed.

The dynamic demand model is estimated using the simulated method of moments. To fit the structural parameters I use an adaptive Markov chain Monte Carlo (MCMC)

[9]This is because the iterative application of the Bellman equation is not necessarily a contraction mapping on an arbitrarily chosen Euclidean norm.

[10]These methods are well suited to sparse high dimensional least squares problem.

method from statistical computing (Chernozhukov and Hong (2003); Łącki and Miasojedow (2015); Baragatti et al. (2013)).

As noted by Imai et al. (2009) and Norets (2009), solving the dynamic demand model at every parameter guess is costly for MCMC methods. In line with the approach suggested by Imai et al. (2009), I estimate the model by alternating between fitting the structural parameters using a single iteration of the estimation algorithm and solving the dynamic demand model using a single iteration of the ADP algorithm.[11]

The model is then applied to the UK laundry detergent industry using household level purchase data from Kantar Worldpanel. The data spans the period from 1st January 2009 until 31st December 2011. For the application, I focus on households who make the vast majority of their purchases at one store - a leading grocery retailer in the UK.

I show that the model closely matches the distribution of brand and format shares. Specifically, by introducing unobserved persistent taste heterogeneity the model closely matches the high level of brand and format loyalty observed in the purchase data.

Further, the model captures key price and inventory dynamics which I show using two policy experiments. These price experiments highlight the role of purchase acceleration from promotional prices, suggest that inventory costs confer market power on manufacturers, and show the importance of unobserved heterogeneity in preferences in understanding consumer dynamics.

The remainder of the paper is structured as follows. In Section 2 I provide an overview of the UK laundry detergent industry and highlight the key economic issues. Section 3 describes the dynamic demand model for the UK laundry detergent industry and details the dimension reduction strategies used. Section 4 discusses the identification and estimation of the model, respectively. The results of the empirical application to the UK laundry detergent industry are presented in Section 5 along with policy simulations. Section 6 concludes.

2 UK Laundry Detergent Market

The nature of products sold in the UK laundry detergent industry make the application of existing storable good demand models to it challenging. In addition to the price and inventory dynamics that are a feature of storable good industries, products are highly differentiated. In the UK, the retailer sells eight brands, five different formats in over ten sizes.

Detergents vary in their efficacy; high quality brands are perceived to provide superior cleaning performance to budget detergents. Further, some types detergents are more effective at removing stains (i.e. powder, tablets), while others are better suited to washing delicate clothes and protecting bright colours (i.e. liquid, capsules). In addition, the format of the detergent leads to consumers tend to use and store them differently. Finally, there are many pack sizes - preventing straightforward application of existing methods.

[11] A similar technique is used by Osborne (2017).

The remainder of this section discusses these issues in more depth. The brands, formats and different pack sizes available are documented. Further, reduced form evidence of price and inventory demand dynamics is presented. I begin with brief description of the data used.

2.1 Data

The analysis of the UK laundry detergent industry is based on individual household purchase data sourced from Kantar Worldpanel. Households that take part in the survey scan the barcode of the items they purchase. Using the scanned barcode, the survey records the price and number of packs bought together with the characteristics of the detergent purchased. In addition, the purchase date and store in which the product was bought is also recorded. The purchase data is supplemented by annually updated household demographics and includes details of household composition.

The data used spans the period from 1st January 2009 until 31st December 2011. It focusses on households who make at least 75 percent of their purchases at one store - a leading UK supermarket.[12] Purchases at other retailers are included in the analysis for these households.

The 10th and 90th percentile of grocery spending per equivalent adult are £30 and £70, respectively.[13] Households in the sample contain 1.7 equivalent adults on average and are observed for an average of 135 weeks. A household spends £48 per equivalent adult on average on weekly groceries, of which £2.31 is spent on laundry detergent.

2.2 UK Laundry Detergent

In the UK, laundry detergent is sold in packages called Stock Keeping Units (SKUs). Each SKU contains a single type of detergent and provides a discrete bundle of washes. Physical and quality differences in laundry detergents affect household utility from consumption, storage costs and the price. The number of washes in a SKU determines the consumption quantity that can be serviced without making new purchases.

In this section I describe the physical attributes of laundry detergent sold in the UK. I discuss how they affect the design of the dynamic demand model and the ramifications they have for the applicability of existing approaches.

2.2.1 Brands and formats

The brand and format of a detergent are its defining characteristics of laundry detergent sold in the UK. Other aspects of differentiation include effective temperature range, whether the enzymes are non-biological, scent, and additional stain removal capacity. However, many of these additional aspects of detergent differentiation are closely tied to

[12] Spending in UK supermarkets accounts for approximately 2 in every 5 pounds of the UK's retail expenditure sector.

[13] The modified equivalent adults scale is used. The first the adult counts 1, then for anyone else over the age of fourteen add 0.5. For those under the age of 14 add 0.3.

the detergent's format and/or only included in high-end brands. Therefore, by focusing on brand and format of a detergent, these other characteristics are effectively captured.

Brands There are six major brands sold by two large manufacturers.[14] In addition, the retailer sells its own private label (PL) detergent, as well as several other smaller niche brands.

The distribution of the price per wash for each brand is summarised by box plots in Figure 1. The average price per wash is £0.20 - the same as Brand E. Brand A, Brand B and Brand D are the premium brands and are priced 24%, 14%, and 10% higher than average price per wash. The cheapest is the retailer's PL and is around 30% cheaper that Brand E on average. Budget brands Brand F and Brand C are priced close to the retailer's PL; their average price per wash is 15% and 22% cheaper than Brand E respectively.

In terms of market shares, the bottom panel of Figure 1 shows that the most purchased brand is the retailer's private label with 34% of all washes purchased - suggesting many consumers are price sensitive and elect to purchase the cheapest product. The budget brands are less popular; Brand C and Brand F have 7% and 9% respectively. Brand E, the mid-range brand, is the second most popular with 21% market share. Suggesting that while some consumers focus on price they also like to consume branded products. This is further supported by the fact that the higher quality brands - Brand A, Brand B and Brand D - together account for around 30% of the market.

This highlights that the demand model must also be able to capture consumer heterogeneity over the quality of detergent purchased and allow it to impact on the amount consumed.

Formats Detergent is available in one of five formats: liquid capsules, gel, liquid, powder and tablets. Each format differs in how it is used, its efficacy, the amount of physical storage it occupies, and its ease of storage.

At the point of consumption, households can choose how much liquid, powder or gel to use in a wash. In contrast, capsules and tablets are sold in pre-measured, discrete dosages. Further, the format may also impact on the type of laundry it is being used for. In particular, the presence of bleach in powder makes it especially suitable for removing deep stains, whereas liquid and gel might be better for delicate garments. Formats also differ in the physical amount of material needed for one wash. In particular, the physical amount of liquid based detergents for a single wash is less than solid detergent. Consequently, powder and tablets are likely to take up more physical storage space than other formats per wash. As a result the format being consumed materially affects both consumer utility and storage costs.

In addition, price per wash varies significantly across detergent formats. Figure 1 uses box plots to summarise the observed distribution of price per wash for each format in the data. The convenience and storage flexibility of capsules and tablets appears to command a 10% price premium relative to powder - the second cheapest format on

[14]For confidentiality, brands and manufacturers are anonymised.

average. While performance of gel is broadly similar to liquid, its novelty, combined with its low storage costs and ease of use commands a 25% price per wash premium relative to liquid.

In terms of volume of washes sold, the bottom panel of Figure 1 shows that powder detergents are the most popular (38%), followed by liquid (21%). Tablets account for 17% and capsules and gel account for around 11% of washes sold.

Given these differences, one might expect utility of consumption and storage costs to vary by format. Therefore, it is desirable to allow household utility from consumption and cost of storage to differ by detergent format.

2.2.2 Pack sizes

The distribution of SKU sizes available in the UK laundry detergent industry is shown in Figure 2. It shows that laundry detergent in the UK can be purchased in a wide variety of sizes. The five most popular SKU sizes only account for 55 percent of purchases purchased. To cover 95% of all purchases, in excess of 13 different SKU sizes are needed. As such, the model must be able to incorporate a choice set with a large number of different size choices.

In existing approaches to estimating dynamic demand models of storable goods a dominant SKU size (Erdem et al. (2003)) or a low-dimensional number of SKU sizes (e.g. Hendel and Nevo (2006a)) is used to reduce the dimension of the price state space to help address the curse of dimensionality. For example, Hendel and Nevo (2006a) impose restrictions on the dynamic demand model to split the household's purchase decision into a static brand choice that is conditional on the size of the SKU purchased and a dynamic discrete choice over which size to purchase. Even with these restrictions, the price state space is high dimensional and gives rise to the curse of dimensionality.

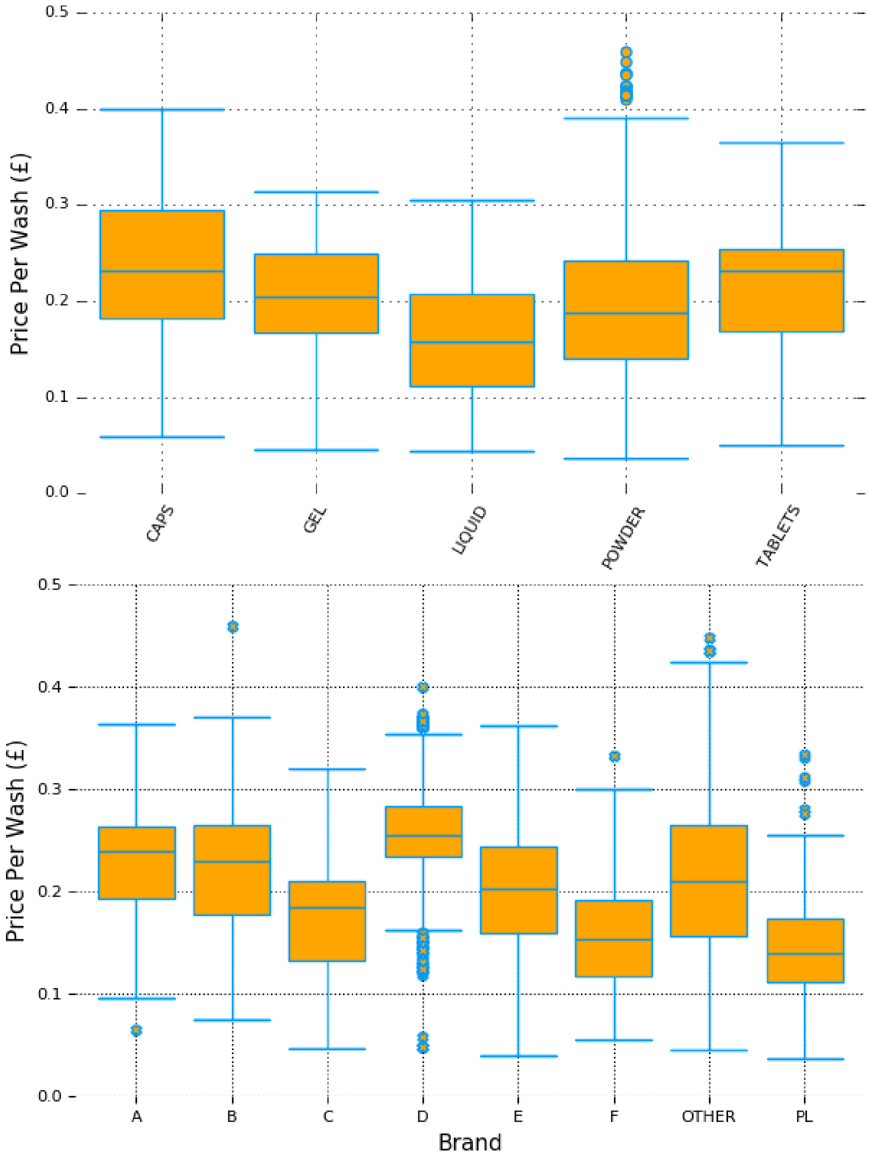
Using the approach pioneered by Melnikov (2000, 2013) and with additional restrictions, the high-dimensional price state space can be replaced by an 'inclusive value' state space with dimension equal to the number of SKU sizes.[15] For example, Hendel and Nevo (2006a) analyse the US liquid detergent market segment between June 1991 to June 1993. In their demand analysis of this market segment there are around ten brands and four sizes. Therefore, these additional restrictions replace a 40 dimensional price state space with an inclusive value state space with four dimensions - one for each size.

In addition to the restrictive assumptions, one drawback of this approach is that it relies on there only being a handful of different SKU sizes. When there are many different SKU sizes, the inclusive value state state is also high dimensional. In this case, this approach does not yield a computationally tractable model. Indeed, there are many markets where the products are sold in many different SKU sizes. The UK laundry detergent is one such industry.

In the spirit of static nested logit models, one might construct groups of ranges of SKU sizes. Then estimate the demand model using a static-brand dynamic-size group split.

[15] 'Inclusive value' refers to the ex-ante expected utility for a household using a random utility model to choose the utility maximising option from a set of alternatives.

Figure 1: Price per wash and market share: by format and brand



Market Shares by number of purchases (%)

Brands	Caps	Gel	Liquid	Powder	Tablets	Total
	Formats					
Brand A	2.87	5.25	0.18	2.03	1.24	11.59
Brand B	2.13	3.31	0.21	3.64	1.22	10.52
Brand C	0.13		0.70	5.51	0.25	6.58
Brand D	0.78	1.14	0.74	3.44	1.13	7.24
Brand E	1.91		8.36	7.22	3.04	20.54
Brand F	0.32		3.64	5.20	0.17	9.33
PL	2.88	1.89	8.82	10.84	9.29	33.72
Other Brands	0.01		0.22	0.08	0.17	0.49
Total	11.03	11.59	22.89	37.96	16.53	100.00

Problematically, conditioning on the size group purchased, only the upper and lower bound of the quantity purchased is known and the inventory available for consumption can take on a range of values. Therefore, consumption, utility, inventory costs, and the next inventory in the next period are set valued. In this case, the dynamic program conditional on purchasing from a group of different size SKUs is not well defined and cannot be solved using standard dynamic programming methods. Hendel and Nevo's (2006) estimation cannot be applied.[16]

It may be tempting to rectify this by using the expected SKU size purchase or another summary statistic of the size group as a size proxy in the state transition function. However, this will necessarily result in a mis-measured inventory and incorrect specification of the transition function of the model.[17]

These approaches runs counter to best practice in the approximate dynamic programming literature developed across a wide range of applications. Powell (2011) states

> "It has been our repeated experience in many industrial applications that it is far more important to capture a high degree of realism in the transition function than it is to produce truly optimal decisions".

This statement reflects the approach taken in this paper. Rather than artificially restrict the households' dynamic choice problem by adding assumptions of convenience, this paper aims to accurately model the high-dimensional price and inventory dynamics. The computational tractability of the model is achieved by combining statistical models of price dynamics and approximates the solution to the model.

2.3 Demand Dynamics

In a dynamic demand model there are two avenues through which demand behaviour is inter-temporally linked. One is through household beliefs about future price movements or promotions. The other is thorough the household's ability to store inventories.

Using observed price and household purchase data, this section assesses whether there is any evidence that both households' beliefs over price expectations and existing inventories create inter-temporal links in demand behaviour.

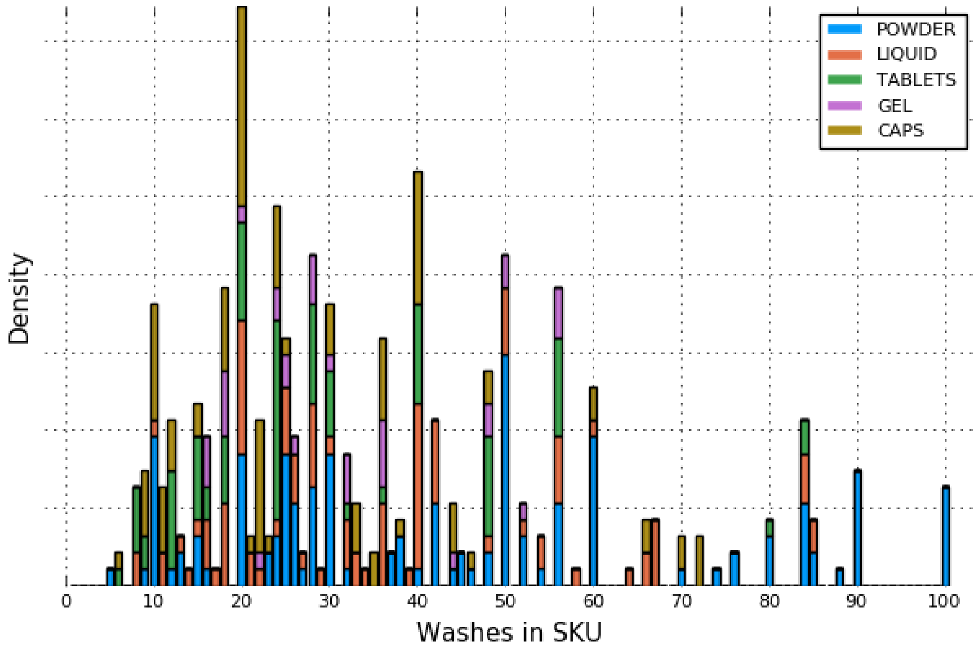
Laundry detergent is often sold on promotion. Figure 3 shows examples of time series of the price per wash for four SKUs that differ in number of washes, brand and format. These price series highlight that the length, depth, and frequency of promotions vary with SKU size, brand and format. Other SKUs exhibit similar price patterns.

The upper left panel of Figure 3 shows that Brand B powder with 42 washes is sold on relatively infrequent deep discounts of around 50% that can last for several weeks. In comparison, the 33 wash Brand D capsules SKU (top-right) discounts are less pronounced and less frequent but typically last longer. Brand E liquid with 28 washes

[16] Osborne (2017) estimates a nested logit model dynamic demand for storable goods by adding assumptions on the ratios of pack sizes within size groups.

[17] These errors might be especially large for size groups with a wide range of sizes.

Figure 2: Sales are not concentrated in a handful of SKU sizes



(bottom-left) is often sold on promotion with around 1/3 off the price - though the discounts vary between 10% to 50%. In contrast, the retailer's private label 24 wash tablets SKU is never promoted. This reflects and everyday low price strategy typical of the retailer's private label products.

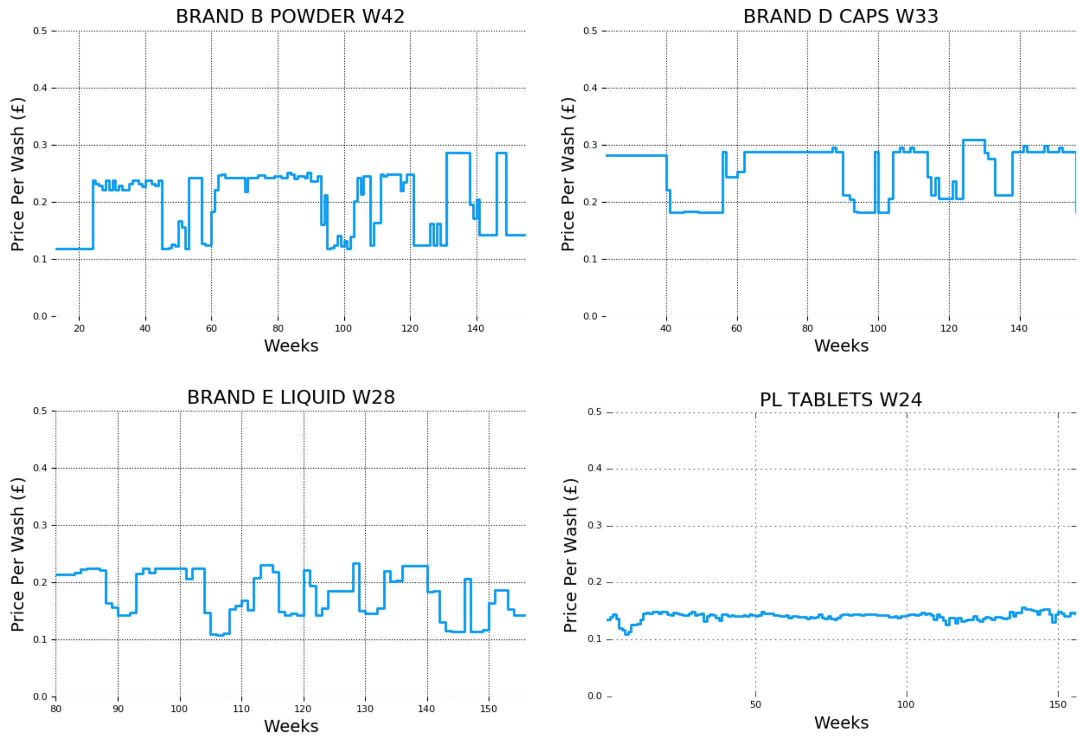
Through repeated shopping trips households are likely to be able to anticipate price movements, and factor them into their purchase decisions. For example, household's may respond to a belief that a promotional price may be short lived by accelerating purchases. Such purchases build up inventories and delay the need to purchase again in the near future. To test for such behaviour, I evaluate the impact of current and past prices per wash on the inter-purchase duration (i.e the number of weeks between purchases_ - a measure of purchase acceleration.

As discussed by Boizot et al. (2001), if price dynamics affect current demand, past and current purchase price per wash should have opposing impacts on inter-purchase duration.

To see why, suppose the previous SKU was an accelerated purchase in response to a sale that the household expected to be short-lived. Following the purchase the household inventories are high. With more detergent in stock, the household can wait longer before they need or want to buy more detergent. Therefore, one should expect inter-purchase duration to be negatively linked to past prices if there is purchase acceleration. Similarly, if a detergent is currently on sale and household's accelerate purchases, the time between purchases shorter. Therefore, one should expect current price price to be positively correlated in weeks since the previous purchase.

Figure 4 shows the result of a regression of the logarithm of weeks since purchase on

Figure 3: Selected price per wash time series



the current and previous logarithm of the price per wash, brand and format together with fixed effects for households. The current price per wash is positively correlated with inter-purchase duration; this suggests that price expectations impact on the timing of purchases are an important aspect of households' purchasing behaviour.

To explore whether the inventories affect the demand for new purchases, figure 5 shows the distribution of inter-purchase duration conditional on the washes per equivalent adult of the previous purchase. It partitions the sample into four groups: households that purchased 0 to 8 W/eq. ad. (top left), 8-12 W/eq.ad. (top right), 12-16 W/eq. ad. (bottom left) and those who bought more than W/eq. ad. (bottom-right).

Figure 5 shows that the distribution of time between purchases shifts rightwards when more washes per equivalent adult are purchased. This is consistent with household's preferring to run down inventories before purchasing again. That is, higher inventories reduce current demand for new purchases.

While the majority of the impact of inventory is expected to be on the decision to purchase, I explore whether current demand conditional on purchase also depends on inventory. To that end, I regress the logarithm of the number of washes purchased on a proxy for inventory, the logarithm price per wash, brand and format dummies, and household fixed effects.[18] The logarithm of price per wash is interacted with an indicator for whether that product was purchased on a sale.

[18]See Annex E for a description of the construction of the inventory proxy.

Figure 4: Conditional correlation of current and previous price per wash with inter-purchase duration

Variable	Estimate	Std. Err.
Current purchase: ln(Price Per Wash)	0.085	0.027
Last purchase: ln(Price Per Wash)	-0.380	0.027
Num. Obs.	11,592	

Note: Includes HH fixed effects and controls for current and previous brand and format purchased

Figure 5: Inter-purchase duration given washes purchase per equivalent adult

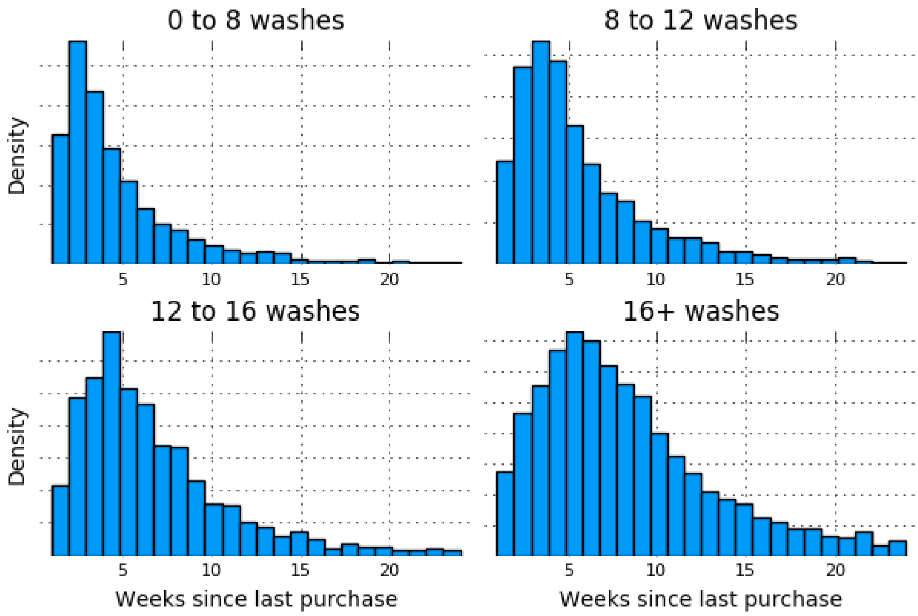


Figure 6: Quantity purchased, inventory, prices and sales

Variable	Estimate	Std. Err.
Inventory/100	-0.105	0.019
Price Per Wash: Regular Price	-0.798	0.023
Price Per Wash: On Sale	-0.813	0.021
Num. Obs.	4,421	

Note: Includes fixed effects for detergents and household

The results, shown in Figure 6, are also consistent with stockpiling behaviour. Namely, higher inventory tends to reduce current demand, promotional activity marginally likelihood of purchasing more washes (i.e. purchase acceleration), and households' demand is inversely linked to price per wash.

Overall, the reduced form analysis of household purchasing behaviour presented in this section supports the view that a static demand model would be mis-specified for the UK laundry detergent industry: both inventories and lagged prices affect demand.

2.4 Persistent taste heterogeneity

Incorporating unobserved taste heterogeneity in static demand models allows for rich substitution patterns. In a dynamic demand model this is especially important because taste heterogeneity leads to persistence in choices and interacts with the timing of households' purchases.

For example, households loyal to a particular brand might accelerate their purchase to take advantage of a promotion of their favoured detergent that they perceive to be short-lived. Alternatively, they may choose to purchase "stop-gap" detergents that share desirable aspects of their preferred choice. In turn, enabling them to continue to smooth their consumption without unduly increasing the cost of purchasing their favoured detergent on sale in the near future.

Figure 7 shows the share of purchases in the current period conditional on the most recent purchase. Without brand and/or format loyalty, one would expect to see similar distributions of purchase shares irrespective of the last format or brand purchased. However, the top panel in Figure 7 shows that households buy the same format on approximately 80% to 90% of purchase occasions. This suggests that there is persistence in preferences of the format purchased.

Similarly, households also exhibit brand loyalty. The bottom panel of Figure 7 shows that brand loyalty is highest at opposite ends of the price distributions. In particular, 86% of households whose last purchase was the retailer's private label buy it on the next purchase occasion. Among the branded detergents the of the higher quality products, Brand A and Brand D, have the highest customer retention; 73% and 79% respectively. Whereas Brand C and Brand F, the 'budget' brands, are re-purchased on 55% and 63% of purchase occasions, respectively. The mid-range brands, Brand E and Brand B, are consecutively purchased 68% of purchase occasions.

Further, if they do switch, they may switch to products that are used and stored in a similar way. For example, those who tend to buy capsules switch most often to tablets and vice versa. Similarly, gel purchasers switch most often to liquid and liquid to gel. As noted earlier, these pairs of formats are the most similar to one another to use.

For brands, Brand E is a popular second choice - especially for consumers switching away from higher quality brands like Brand A, Brand B and Brand E. Similarly, the retailer's private label is also popular - especially when switching away from lower end brands such as Brand E and Brand F.

This analysis shows that households tend to purchase the same brand and format. In turn, suggesting that persistent taste heterogeneity is an important component of

Figure 7: Conditional purchase shares: format and brand purchased

			Current Format		
Last Format	Caps	Gel	Liquid	Powder	Tablets
Caps	82.0	3.1	5.1	3.2	6.6
Gel	3.1	78.8	13.5	3.5	1.0
Liquid	3.0	8.1	81.4	5.5	2.0
Powder	1.3	1.3	3.5	91.8	2.0
Tablets	6.1	1.6	3.2	4.3	84.7

				Current Brand			
Last Brand	Brand A	Brand B	Brand C	Brand D	Brand E	Brand F	PL
Brand A	72.8	6.2	2.1	1.9	8.7	2.9	5.3
Brand B	7.0	67.5	3.2	1.1	7.0	8.2	6.1
Brand C	4.2	6.5	62.3	0.6	6.7	12	7.5
Brand D	3.5	1.7	0.7	79.2	8.5	1.0	4.7
Brand E	5.3	4.8	2.1	4.2	67.5	5.3	9.9
Brand F	4.2	9.2	8.7	0.9	11.4	54.8	10.3
PL	1.6	2.4	1.6	0.8	4.6	2.3	86.0

household's purchase decisions.

3 Dynamic Demand Model

Given inventories, prices, and income, households choose purchases and consumption to maximise the sum of current and expected future utility net of inventory costs and shopping costs.

In the dynamic demand model each household visits the supermarket in each week. On each visit, a household decides whether or not to purchase laundry detergent. If they purchase, a household chooses the format, brand and pack size from a set of around 100 alternatives.

When making purchases households take into account their preferences, their inventories, current prices and their beliefs of the evolution of future prices.

From a household's viewpoint, the timing, length and depth of the discounts of price promotions are uncertain. In the face of this price uncertainty, I assume households use a statistical model to forecast prices. Because they frequently shop at the same supermarket, they observe a long history of prices they can use to build such a forecasting model. Recognising that tracking and forecasting 100 or so prices is cognitively challenging, I assume that households use a low rank approximation of the underlying price movements. Specifically, households use a hedonic model to capture cross sectional variation in SKU prices and a dynamic factor model to predict price changes over time. This model is parsimonious and captures the salient features of the dynamic pricing environment.

Purchases are added to stocks held at home. Then after purchasing, households choose how much to consume of each of the laundry detergents available at home. For example, a household may choose to use high quality liquid detergent for delicate clothing and lower quality powder detergent for washing other clothes.

Any detergent left over from after consumption is retained to inventories and storage costs are incurred. Households use valuable space in their homes to store laundry detergent.[19] Because detergent competes with other products for limited storage space, the marginal cost of storing another wash is (weakly) increasing. Moreover, the ease of storing different formats of detergent is reflected in marginal inventory costs.

With convex storage costs the increment in inventory costs for purchases is minimised when stocks are low. As a result, household may elect to purchase infrequently to restock laundry detergent - a feature of observed purchase sequences in the data.

In the remainder of this section, I first detail current period utility and the budget constraint, then inventory and shopping costs and the transition law for inventories, and finally beliefs about future prices.

3.1 Household utility

Households are infinitely lived and discount the future at a rate $\delta \in (0,1)$. In each period, households get utility from consuming detergents and a composite of all other products. Let the utility function in each period t for household i be

$$U_{it} = U(C_{it}) + \psi_{i0} C_{i0t} \tag{1}$$

where $C_{it} = [C_{i1t}, \ldots, C_{iJt}]^\top$ and C_{ijt} is the amount of detergent j consumed by household i in period t. The quantity of the composite good consumed is C_{i0t} and ψ_{i0} is the marginal utility from consuming it. To focus on the demand dynamics in the laundry detergent market, the composite good is assumed to be non-storable.

The per period utility is assumed to be additively separable in the utility from consuming laundry detergent and the utility from consuming the composite product. This rules out any complementaries between other products and laundry detergents.

The utility from consumption of laundry detergent is assumed to be quadratic.

$$U(C_{it}) = \psi_i^\top C_{it} - \frac{1}{2} C_{it}^\top \Psi_i C_{it} \tag{2}$$

It has two sets of parameters for household i: ψ_i, Ψ_i. The first set, ψ_i, is a J-vector whose j-th element, ψ_{ij}, is the marginal utility of the first unit consumed of detergent j for household i. The second set, Ψ_i, control the rate at which detergent is consumed. The j-th diagonal element of Ψ_i governs the rate at which marginal utility

[19] In 2016 the UK average purchase of housing per square foot is around £200. Using the average rent-to-house price ratio, this translates into around £1.20 per square foot per month.

of consumption of detergent j declines as more of the detergent is consumed.[20] The off-diagonal elements of Ψ_i allow marginal utility of consumption to depend on the amount of other detergents consumed. These terms affect the degree to which detergents are substitutes (or complements). To ensure the utility from consuming detergent is concave in consumption, I assume Ψ_i is a $J \times J$ symmetric positive definite matrix. If all elements of Ψ_i are zero, then the utility function collapses to the linear specification in Erdem et al. (2003).[21]

Both ψ_i and Ψ_i allow for persistent taste heterogeneity over consumption of different detergents. As noted in Section 2, detergents vary in terms of their quality and ease of use. Moreover, Section 2.4 shows that households repeatedly purchase the same brand and/or format of detergent. This indicates that both brand and format loyalty are important determinants of demand and highlight the importance of including persistent taste heterogeneity in the model.

3.2 Purchases and inventory

Prior to choosing how much to consume, a household decides which, if any, detergent SKU to buy. Detergents are sold in M distinct SKUs indexed by $m = 1, \ldots, M$. As noted in the description of the UK laundry detergent industry in Section 2, the number of SKUs on sale in the retailer in each week is around 100 and there are 37 types of detergent available for purchase over the sample period (i.e. $M > J$).

Let $d_{imt} = 1$ if household i purchases SKU m in period t and is zero otherwise. Adopting the convention that $m = 0$ indexes a household's decision not to purchase, $d_{i0t} = 1$ when the household i does not buy detergent in period t. Households are restricted to a single purchase in each period, $\sum_{m=0}^{M} d_{imt} = 1$.

To capture utility from other SKU related factors observed by the household, but not by the econometrician, identically and independently distributed SKU specific utility shocks $\varepsilon_{imt} \overset{iid}{\sim}$ Type I Extreme Value with scale parameter σ_ε are added to the per-period utility function (eq (1)).

In each period, household's per period income is Y_{it}. The amount spent on SKU purchases by household i in period t is $P_{it} = \sum_{m=0}^{M} d_{imt} P_{mt}$ where P_{mt} is the price of SKU m in period t. In addition, households also incur shopping costs, PC_{it}, and inventory costs, IC_{it}. The remainder of the income is spent on the quantity consumed of the composite good whose price is normalised to 1.[22] The the budget constraint is

$$Y_{it} = C_{i0t} + P_{it} + PC_{it} + IC_{it} \tag{3}$$

[20] To see this consider a static model with this utility function and a single good. In this case, the interior optimal consumption is given by $C_{ij} = \psi_{ij} / \Psi_{i,(j,j)}$.

[21] This specification also captures the key qualitative features of the utility function in Hendel and Nevo (2006a). In their paper all detergents are perfect substitutes at the point of consumption and the utility function is strictly concave. This occurs when all detergents have the same linear utility parameter, the same non-zero value of the diagonal component of Ψ_i, and the off-diagonal terms of Ψ_i are constant and identical.

[22] In this case, ψ_{i0} is also the marginal utility of income.

Substituting the budget constraint into the per-period utility function, yields the flow payoff for households after having made purchases and consumed detergent,

$$U_{it} = U(C_{it}) + \psi_{i0}\left(Y_{it} - P_{it} - PC_{it} - IC_{it}\right) + \frac{1}{\sigma_\varepsilon}\sum_{m=0}^{M} d_{imt}\varepsilon_{imt} \tag{4}$$

Next, shopping costs and inventory are discussed in more detail.

3.2.1 Shopping Costs

Shopping costs include the time, search and carrying costs of purchasing a SKU. Shopping costs are given by

$$PC_{it} = \begin{cases} \rho_0 & \text{if } d_{i0t} = 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

When shopping costs are high, households prefer to make as few purchases as possible. However, they have no bearing on the timing of purchases - in the model inventory costs is the key determinant of inter-purchase duration. I return to this in more detail when inventory costs are discussed in Section 3.2.2.

3.2.2 Inventories

Purchases are added to existing inventories, $I_{it} = \left[I_{i1t}, \ldots, I_{iJt}\right]^\top$. The post-purchase inventory is,

$$\bar{I}_{it} = I_{it} + Q_{it} \tag{6}$$

where Q_{it} is a J-vector whose j-th element is $q_{j,m}$ - the number of washes of detergent j available for consumption when SKU m is purchased. All other elements of Q_{it} are zero. Further let $\bar{J}_{it}$ denote the number of distinct detergents held in inventory by household i at time t.

After purchasing, households choose how much of the inventory held in stock, if any, to consume. There are four possibilities. First, they can choose to consume some, but not all of the detergent. Second, the remaining stock of the detergent can be consumed (i.e. only a small amount is left). Third, they can elect not to consume any, even if it is held in stock. The final case is that there is no detergent consumed or held in stock.

Unused inventory, $I_{it+1} = \bar{I}_{it} - C_{it}$ comprising of $\bar{J}_{it+1} = \sum_{j=1}^{J} \mathbf{1}\left[\bar{I}_{ijt} > 0\right]$ distinct detergents are held in inventory at the end of the period. They are stored for future consumption at an inventory cost to household i in period t of

$$IC_{it} = \gamma_i^\top I_{it+1} + \frac{1}{2} I_{it+1}^\top \Gamma_i I_{it+1} + \kappa_1\left(\bar{J}_{it+1} - 1\right) + \kappa_2\left(\bar{J}_{it+1} - 1\right)^2 \tag{7}$$

where γ_i and Γ_i are a J-vector and a $J \times J$ symmetric matrix of inventory cost parameters, respectively. The maximum number of detergent stocked at any time is assumed to be three.[23]

Inventory costs depend on (i) the amount of washes, and (ii) on the number of different types of detergent in inventories at the end of the period.

The total amount of washes are an important determinant of inter-purchase duration. With convex inventory costs, the incremental storage costs of new purchases are lowest when existing stocks are low. As a result, households prefer to run down detergent stocks before repurchasing. Provided that purchased quantities are large relative to weekly consumption, a household would choose to infrequently restock laundry detergent - a feature of the observed purchase sequences in the data.

The inventory costs associated with stocking different types of detergents can capture the fact that SKUs are bulky and tend not be discarded until they are used up.[24] Assuming that the costs of stocking distinct SKUs are increasing for one or more SKUs, holding nine washes of one detergent is less costly that holding three washes of three different detergents.

An alternative interpretation is that costs associated with stocking different detergents is that households incur switching costs. Combined with the fact that households prefer to purchase before stocks run out, the costs associated with storing different types of detergents act as brand-format switching costs. Therefore, they help explain the high repurchase rates observed in the brand and format purchase transition matrices in Figure 7.

3.3 Household price forecasts

The timing of purchases is also affected by expected price movements. As highlighted in Section 2, laundry detergent is often available for purchase on promotion. From the perspective of the household, the depth of the price discount, the timing, and the length of the promotion are uncertain.

The possibility of a change in the price in the near future will affect the current purchase decision. For example, suppose that a SKU is available on a deep discount

[23]The cost of stocking four or more different detergents is assumed to be infinite. A simulation of inventories using observed sequences of 620 households in the Kantar Worldpanel data suggests that household stock less than three different detergents in over 99.5% of all simulated periods. In this simulation household are assumed to consume detergent on a first-in-first-out basis. The target quantity consumed is equal to the total number of washes the household purchases divided by the number of weeks they are observed in the data. In each period, households consume the smaller of the total number of washes held in inventory and the target consumption for that household. The simulation assumes that households initially have zero inventories.

[24]With an additional assumption on the order in which SKUs containing the same detergent are used up (i.e. first-in-first out), the number of SKUs held at any one time can be tracked. The simulation described in footnote 23 with the first-in-first out assumption suggests that households stock three of fewer SKUs in around 95% of simulated periods. The maximum number of SKUs stocked in the simulation is 12, but in 99% of the periods in the simulation household stock five or fewer SKUs. The number of SKUs in stock is quite similar to the number of distinct detergents held. This suggests that household prefer to run down detergent in individual SKUs before restocking. Therefore, the detergent specific stocking inventory cost in equation (7) may be a good approximation to the more complicated inventory transitions in which the number of SKUs are recorded.

that the household believes to be short lived. Even though they may have enough inventory to service consumption needs for the near future, they may accelerate the SKU purchase to take advantage of the relatively low purchase price.

To forecast prices, households build a forecasting model using a long history of prices observed on repeatedly making shopping trips to the same supermarket. Households are assumed to use a statistical model to forecast prices based on prices they have observed. In particular, households use a low-order Markov process to forecast M SKU prices in the next period, $P_{t+1} = [P_{1,t+1}, \dots, P_{M,t+1}]^{\top}$,

$$P_{t+1} \sim G_{P_{t+1}|P_{t,\tau}} \tag{8}$$

where $P_{t,\tau} = [P_t, P_{t-1}, \dots, P_{t+1-\tau}]$ is the matrix of past τ-periods prices.

Recognising that tracking and forecasting 100 or so prices is cognitively challenging, households use a low rank approximation of the underlying price movements.

The household's price forecasting model has a hedonic component that captures variation prices across brands and formats and allows for quantity discounting. The variation in price due to promotions is modelled using a dynamic factor model. The forecasting model is estimated using observed prices and its dimensions and parameters are chosen using statistical criteria. Further details are deferred until Section 3.5.1.

Next, I show that the household's dynamic problem can be written as a two-stage discrete continuous choice problem.

3.4 Household choice problem

Bringing together the elements of the model described above, the household's choice problem can be written as a two stage discrete-continuous Markov decision problem expressed in recursive form.

In each period, each household faces a discrete-continuous decision: (i) they choose whether or not to purchase laundry detergent, and (ii) how much laundry detergent to consume in the current period.

To describe how households choose which detergent tp purchase and how much to consume, it is instructive to work backwards and start with the consumption decision.

As in Dubin and McFadden (1984), households first solve for optimal consumption conditional on choosing option m from the choice set. That is, the consumption quantity of each detergent held in stock after purchasing is given by the solution to the Bellman equation

$$\begin{aligned} W\left(s_{it}\right) &= \max_{0 \le C_{it} \le \bar{I}_{it}} U(C_{it}) + \psi_{i0}\left(\bar{Y}_{it} - PC_{it} - IC_{it}\right) \\ &\quad + \delta \int \ln \sum_{m=0}^{M} \exp\left\{W\left(s_{it+1}\right)\right\} dG_{P_{t+1}|P_{t:\tau}} \end{aligned} \tag{9}$$

where $s_{it} = [\bar{I}_{it}, \bar{J}_{it}, P_{it}, \bar{Y}_{it}]$ is the state space after purchases have been made in period t containing the post-purchase inventories, $\bar{I}_{it}$, the number of distinct detergents in inventory, $\bar{J}_{it}$, prices, P_t, and income net of the SKU price paid, $\bar{Y}_{it} = Y_{it} - P_{it}$.

Then, households choose which, if any, SKU to purchase from the choice set by maximising the sum of utility shocks and the conditional indirect utility function,

$$\begin{aligned} V(s_{it}, \varepsilon_{it}) &= \max_{\{d_{imt}\}_{m=0,\ldots,M}} W(s_{it}) + \frac{1}{\sigma_\varepsilon}\varepsilon_{imt} \\ s.t. \quad & \sum_{m=0}^{M} d_{imt} = 1 \\ & \varepsilon_{imt} \overset{iid}{\sim} \text{Type I EV} \end{aligned} \tag{10}$$

Written as a two-stage problem, the solution of the model requires a search for a conditional indirect utility function and policy for consumption to solve equation (9).[25]

As specified, the model has a very large state space. The vector s_{it} includes 37 inventory variables - one for each detergent - and in excess of 100 price variables. The curse of dimensionality bites hard and renders the problem computationally intractable. The next section discusses this issue in more detail and outlines the approach taken in this paper that makes the model computationally estimable.

3.5 The curse of dimensionality

As highlighted in Section 2, there are many different laundry detergents a household can purchase and store. Consequently, the inventory and price components of the state space are high-dimensional continuous variables. This acutely exacerbates the curse of dimensionality.

This paper proposes an approach to alleviate the curse of dimensionality. Unlike existing approaches it can easily be adapted for application to other storable good industries. Central to the dimension reduction strategies is to address the different ways in which the curse of dimensionality arises from the information needed to forecast future prices and the number of different detergents that a household can store.

The proposed approaches to alleviate the curse of dimensionality that arise from price and inventories are discussed in turn. Subsequently, the household's choice problem is revisited.

3.5.1 Forecasting SKU Prices

Through repeated visits to the store to purchase laundry detergent, households observe a long sequence of prices they can use to forecast prices.

Households are assumed to forecast the price per wash for each SKU, $p_{m,t}$. Then SKU prices are forecasted by multiplying the predicted price per wash for SKU m by the number of washes contained in each SKU. That is, the expected SKU price in period $t+1$ is,

[25] Annex B derives the two-stage representation of the problem.

$$\mathbb{E}_t\left[P_{m,t+1}\right] = \mathbb{E}_t\left[p_{m,t+1}\right] q_{j,m} \tag{11}$$

where $q_{j,m}$ is the number of washes of detergent j contained in SKU m.

As highlighted in Section 2.3, the average price per wash varies across brands and formats. Further, there are quantity discounts. In addition to differences in the average price per wash across SKUs, SKU prices vary over time because they are often sold on promotion. Figure 1 shows that the length, depth, and frequency of promotions vary with SKU size, brand and format.

When forecasting prices, households predict both the cross-sectional and time series components of the price per wash across different SKUs. To forecast the average price per wash across SKUs, households are assumed to use a hedonic approach. Specifically, they model average price per wash as a function of the brand, format and controls for the number of washes in the SKU.

Tracking the promotional price activities of over 100 products is likely to be cognitively prohibitive. Instead, I assume that households use the history of prices observed from repeated shopping visits to formulate a boundedly rational price forecasting model.

While households are unlikely to be able to track all SKU prices, I assume that they are able to keep track of a low dimensional set of price trends that underpin price dynamics. Further, I assume that they understand how the promotions of individual SKUs interact with these underlying price trends. For example, they know that the retailer's private label SKUs are almost never promoted. Whereas, large SKUs containing high quality branded detergents are known to be heavily but infrequently discounted. They also expect that major brands tend to be promoted asynchronously, rather than at the same time.

In line with this bounded-rationality approach to forecasting prices, I assume households use a factor model as a low-rank approximation of the time series component of the price forecasting problem. Bringing this together with the hedonic model of cross sectional price variation, households' forecasting model for SKU prices is an interactive fixed effects model (Bai (2009))

$$z\left(p_{mt}\right) \quad = \quad \lambda_m^\top F_t + X_m^\top \alpha + \epsilon_{imt} \tag{12}$$

for $t = 1, \ldots, T$ periods observed in the data. F_t is an R-vector of the price factors in period t, λ_m is an R-vector of factor loadings for SKU m, and X_m are the SKU characteristics used in the hedonic model to capture cross-sectional price variation.[26] The dependent variable forecast price per wash is $z(p_{mt}) := \frac{1}{2}\left(1 + e^{-p_{mt}}\right)^{-1}$. This transformation prevents the prediction of negative prices and provides a sensible upper bound motivated by observed data.[27]

The SKU price movements are governed by the dynamics of price factors, F_t. I assume factors follow a stationary, τ-order exogenous Markov process,

[26]Scale and rotation normalisation restrictions are imposed on price factors and factor loadings. See Stock and Watson (2002); Bai and Ng (2002); Bai (2009) for further details.

[27]All observed price per wash lie in the interval £0.00 and £0.50

$$F_t = A_0 + \sum_{s=1}^{\tau} A_s F_{t-s} + u_t \tag{13}$$

where $\mathbb{E}\left[\epsilon_{imt}|F_t, X_m\right] = 0$, $u_t \perp \epsilon_{imt}|F_{t,\tau}, X_m$, $F_{t,\tau} = [F_t, F_{t-1}, \ldots F_{t-\tau+1}]^\top$ and $u_t \overset{i.i.d.}{\sim} N(0, \Sigma_u)$. The number of lags included τ is informed by statistical criteria and relies on data.

The resulting dimension of the price factors in the state vector is τR. When prices are correlated over time, the price factor state space, F_t, will likely contain only a few dimensions. In many applications the resulting reduction in the size of the state space is substantial. As discussed in detail in Section 5.1, $\tau = 1$ and $R = 2$ in this paper.

Beneficially, this approach does not rely on the number of brands or pack sizes sold and can be applied to any industry. Moreover, rather than add restrictive ad hoc assumptions, the curse of dimensionality is substantially mitigated using a statistical criterion that relies on observed price data.

3.5.2 Inventories

Where the inventory dimension contains more than a handful of products, the existing literature has sought to reduce the dimension of inventory by imposing restrictions on the household's problem. However, rather than add assumptions, an approximate dynamic programming (ADP) approach is used.

The defining feature of an ADP method is that it retains the high dimensional representation of the household's choice problem but seeks a high quality but computational feasible approximate solution. Indeed, the approximate solution to this complex choice problem can be viewed as the result of a boundedly rational household.

In this paper, there are two components to the approximation to $W(s)$: the approximation architecture and the selection of a set of 'features' of the high-dimensional problem that characterise the dimension reduction. In economics the approximation architecture usually describes the function space in which the approximate solution is sought. Feature selection is analogous to model selection; it is the method or criterion function that is used to select elements from this function space. Below I describe these ideas using in the language of approximate dynamic programming.

Approximation architecture Adopting Powell's (2011) terminology, there are three broad categories of value function approximation architectures: a lookup table approximation (i.e. solving the dynamic program on a fixed grid), a linear-in-parameter approximation (i.e. polynomials, B-splines, etc), and a non-linear approximation (i.e. neural nets). In this paper, I focus on linear-in-parameter approximations

$$W(s_{it}) \approx \phi(s_{it})^\top r \tag{14}$$

where $\phi(s_{it})$ is an L-vector of basis functions and r is an L-vector of parameters.

The reasons for choosing a linear-in-parameters approximation are two-fold. First, lookup tables tend to suffer from the curse of dimensionality for even moderately sized problems. In this case, even with an aggressive dimension reduction strategy, it is likely that the model could only be solved on a very coarse discretisation of the state space.[28]

Second, linear-in-parameters approximations can exploit the linear structure to ensure updating is relatively straightforward and are easily modified to address numerical stability issues. Whereas, non-linear architectures are more complex to update and less numerically stable. As a result, there are no convergence guarantees and *sui generis* modifications are needed to improve performance.[29]

To approximate $W(s_{it})$ a class of flexible polynomials suitable for large-scale dynamic programming problems called Smolyak polynomials is used.[30] Beneficially, the number of basis functions used by this class of models grows polynomially, rather than exponentially as the state space expands. Moreover, if an anisotrophic Smolyak polynomial is used, the accuracy of the approximation can be varied in each dimension of the state space.[31]

Taken together, the linearity, smoothness and sparseness of the approximation architecture allow for a richer set of low-dimensional features to be included in the approximation without incurring too high a computational penalty.

Feature selection In ADP, a high quality, low dimensional approximation to the solution is created by identifying and extracting the salient features of the problem. Typically this involves some transformation or aggregation of the high dimensional state space and exploits the structure in the dynamic program.

In retail storable good industries, like laundry detergent, households are unlikely to stock more than a handful of products at once. As a result, the inventory vector is likely to be sparse. This sparsity is the feature of the household's choice problem used to reduce the dimensionality of the inventory vector in this class of models.

The inventory state space entering the approximation is restricted to include only those detergents that are held in stock after purchases. The subset of coefficients of r that multiply all basis functions that depend on the inventory of other types of detergent are set to zero.

To illustrate how this works in practice, consider a household that has three washes of Brand C powder in stock at the beginning of period. If they make no purchases, only basis functions that depend on the inventory of Brand C powder and other non-inventory

[28] Nevertheless, other more sophisticated approximations that work with discretised high-dimensional state spaces, such as hierarchical approximations (see Powell (2011); Bertsekas (2011a))) and adaptive grid methods (Brumm and Scheidegger (2015)) may prove to be fruitful.

[29] See Powell (2011) for a discussion of updating nonlinear approximation architectures. Note that neural nets can be easily updated in some instances. As highlighted by Bertsekas (2011a); Judd et al. (2011), an additional benefit of using a linear architecture is that other projections can be considered (i.e. regularisation can be added).

[30] See Malin et al. (2011); Judd et al. (2014); Maliar and Maliar (2014) for an overview of this class of polynomials and a detailed discussion of how to efficiently implement them.

[31] The polynomial growth of basis functions involving each variable is linked to the chosen level of accuracy.

state variables are included in the approximation to $W(s_{it})$. All other basis functions are excluded by setting their coefficients zero.

Suppose now that the household purchases 20 washes of Brand A Gel. In this case, all of the basis functions that depend on inventories of either Brand C powder or Brand A gel and other non-inventory state variables are included in the approximation to $W(s_{it})$. Again, the coefficients on basis functions that depend on inventories not stocked are set to zero.

The benefit of reducing the number of inventory state variables used to approximate $W(s_{it})$ is that the number of basis functions that need to be evaluated at each state is greatly reduced. As a result, the the computational burden of approximating the solution to the household's choice problem is significantly lowered.

Formally, to implement this approximation strategy define matrix Ω_{it} to select the subset of basis functions select that depend directly on the detergent held in stock and other non-inventory state variables. The lower-dimensional subset of basis functions and corresponding parameters are calculated by pre-multiplying the high-dimensional counterparts by Ω_{it},

$$\tilde{\phi}(s_{it}) := \Omega_{it}\phi(s_{it}) \tag{15}$$

$$\tilde{r} := \Omega_{it}r \tag{16}$$

The approximate conditional indirect utility function is

$$W(s_{it}) \approx \phi(s_{it})^{\top} \Omega_{it}^{\top}\Omega_{it}r = \tilde{\phi}(s_{it})^{\top} \tilde{r} \tag{17}$$

Even though the dimension of the approximation for a particular inventory configuration is lowered, this dimension reduction strategy for inventory does not reduce the number of parameters. That is, I need to solve for the all of the parameters in r. While this may seem to be an obstacle, compared to the cost of forming the approximation to $W(s_{it})$ defined on even moderately sized state spaces, it is relatively cheap to fit r.

This is because parameters are only updated once using stochastic projected gradient descent methods in each iteration of the ADP algorithm.[32] These computationally light recursive methods are well suited to sparse, high dimensional optimisation problems. Therefore, fitting many parameters is not a computationally expensive part of the algorithm.

3.6 Solving the ADP

Incorporating the household's price forecasting model into the choice problem, households need only keep track of a coarser partition of the price state space containing τR price factors to forecast SKU prices. The conditional indirect utility

[32]See Bertsekas (1999) and Parikh et al. (2014) for further details of stochastic projected gradient descent methods.

function is redefined on the coarser partition of the price state space, $s_{it} = \left[\bar{I}_{it}, \bar{J}_{it}, F_t, \bar{Y}_{it}\right]$.[33]

Implementing the dimension reduction strategy for inventories, the approximation to the solution of the household's choice problem is constrained to lie in the function space, $\mathcal{S}^{\phi} := \left\{\phi\left(s_{it}\right)^{\top} \Omega_{it}^{\top}\Omega_{it} r | r \in \mathbb{R}^L\right\}$.

However, the conditional indirect utility function that solves the Bellman equation (eq. (9)) may not lie in the function space $\mathcal{S}^{\phi}$. If not, there is no $r \in \mathbb{R}^L$ that solves the household's optimal consumption problem and we must look for an approximate solution.

One alternative is to project $W(s)$ onto S^{ϕ} and find the value of r that minimises the distance between $W(s_{it})$ and its approximation with respect to some norm. To implement this, r is chosen to minimise the Bellman equation approximation errors with respect to a Euclidean norm, $\|\cdot\|$

$$r^{\star} = \arg\min_{r \in \mathbb{R}^L} \|\Phi r - T(\Phi r)\|^2 \tag{18}$$

where $T(\Phi r)$ is the Bellman optimality operator in eq. (9) and

$$\underset{(N \times L)}{\Phi} = \begin{bmatrix} \phi\left(s_1\right)^{\top} \Omega_1^{\top}\Omega_1 \\ \vdots \\ \phi\left(s_N\right)^{\top} \Omega_N^{\top}\Omega_N \end{bmatrix} \tag{19}$$

is a sparse $N \times L$ matrix of basis functions evaluated at $n = 1, \ldots, N$ state space vectors.

Equivalently, $r^{\star}$ can be expressed as the fixed point of the projected Bellman equation,

$$\Phi r^{\star} = \Pi T\left(\Phi r^{\star}\right) \tag{20}$$

where Π is the projection with respect the Euclidean norm, $\|\cdot\|$. To find $r^{\star}$ that defines the fixed point of equation (20), I use an ADP algorithm called λ-policy iteration (Bertsekas (2015)) - an algorithm designed to solve for points of a projected Bellman equation.[34]

Like exact policy iteration, this algorithm is iterative. At each of the $k = 0, 1, \ldots, K$ iterations it repeatedly applies two steps: (i) policy evaluation, and (ii) policy improvement.[35] Before, describing these two steps in more detail, I first define the consumption policy function that is evaluated and improved upon in each iteration of the λ-policy iteration algorithm.

[33] To avoid an unwieldy proliferation of function notation, with a small abuse of notation $W(\cdot)$ is redefined on the coarser partition of the price state space.

[34] λ-policy iteration was developed by Bertsekas and Ioffe (1996). The specific algorithm used is called λ-PI(1) and was proposed by Bertsekas (2015), where further details can be found.

[35] As is common in ADP algorithms, a maximum number of iterations is chosen at the beginning of the algorithm. If the algorithm does not converge according to a pre-specified criteria on r, the output of the last iteration is the solution to the approximate dynamic program.

There are two inputs into the consumption policy function. One is the estimate of the solution to eq. (20) at the beginning of iteration k of the λ-policy iteration algorithm, r_k. The other is the state vector, s_n, for household i in period t. Given values for these inputs, the consumption policy produces a J-vector of consumption, $C_{n,k}$, that solves the envelope condition of the household's choice problem

$$\begin{aligned} \nabla\phi\left(s_n\right)^{\top}\Omega_n^{\top}\Omega_n r_k &= \psi_i - \Psi_i C_{n,k} && (21) \\ s.t. \qquad & 0 \leq C_{n,k} \leq \bar{I}_n && (22) \end{aligned}$$

where $\nabla\phi\left(s_n\right)$ is a $L \times J$ Jacobian matrix whose (l, j)-th element is the partial derivatives of l-th basis functions evaluated at s_n with respect to detergent j's inventory.

The benefits of using the envelope condition to define the policy function are two-fold. First, it is computationally light. Specifically, it does require the computational cost of computing expectations. Moreover, because the per-period utility function in this paper is quadratic and concave in consumption, its solution is unique and can be found using a linear equation solver. In the context of an algorithm that heavily utilises simulation, such as λ-policy iteration, this is an important practical consideration.[36]

However, there are also drawbacks of using the envelope condition. Most notably, it is not necessarily a contraction mapping.[37] However, Arellano et al. (2014) show that there exists a damping parameter that can be used to exponentially smooth updates to the value function and ensure that it is contraction mapping. I return to this point in the discussion of the policy improvement step.

Next, I provide further details on how policies are evaluated and improved in the λ-policy iteration algorithm. The steps are discussed from the beginning of iteration k of the algorithm. At the beginning of the iteration, the current guess at the solution to eq. (20), r_k, is plugged into eq. (21) to define the consumption policy function.

In the evaluation step of the λ-policy iteration algorithm, simulation is used to approximate the value of the consumption policy function.

The value of using this policy in a given state is calculated using simulation. The simulation contains $i = 1, \ldots, H$ households. Household i's observed characteristics are uniformly drawn from the sample of households in the Kantar Worldpanel data described at the beginning of Section 2.[38] Persistent unobserved taste heterogeneity and SKU-specific utility shocks for each period are drawn independently for each household and are fixed at the beginning of the algorithm. Observed SKU prices are used in the simulation as are the price factors estimated using eq. (12).

[36] Maliar and Maliar (2013) and Arellano et al. (2014) find that ECM achieves similar speed-ups in computation time over value iteration and the endogenous grid method.

[37] See Arellano et al. (2014) for a detailed exposition. Intuitively, this occurs because ECM does not impose the first order conditions at every step in the iteration, only in the limit. Likewise, value iteration does not impose the envelope condition during the contraction, only in the limit. However, imposing the first order condition is necessary to guarantee that value iteration and other backward iteration methods, like endogenous grid method, have the contraction mapping property.

[38] These include their income, the number of equivalent adults in the household, and the time they first appear in the Kantar Worldpanel data.

The simulation is initiated by setting each household's inventories to zero. Prior to making purchases, each household calculates consumption using the policy under evaluation conditional on purchasing SKU m (or making no purchase).

Then, to decide, which, if any SKU to purchase, each household calculates the conditional indirect utility associated with each purchase option as the sum of the per period utility and discounted continuation value evaluated at r_k.[39] They then pick the option that maximises their conditional indirect utility and add it to inventory. Households then consume a quantity of detergent from inventory determined by the policy being evaluated. Finally, in preparation for the next period, each household's inventory is updated for purchases and consumption decisions. This is repeated for all periods in the simulation.

The simulated value of the policy is calculated once the simulation is completed. To illustrate how, suppose that household i at time period t is in state s_{it} after having made purchases. The simulated value of the policy at state s_{it} is calculated as the net present value of the sum of simulated flow payoffs from t until the end of the simulation and the discounted continuation value in the final period.

To reduce dependence on initial conditions, 12 periods at the beginning of the simulation are excluded from the valuation of the consumption policy. The remaining T_i periods of the simulated sequence of purchases and consumption for household's $i = 1, \dots H$ are used to evaluate the policy. The length of the simulated sequence used to approximate the value function is determined by the parameter λ. Specifically, T_i is drawn from a $Geometric(1 - \lambda)$ distribution.[40]

To make full use of the simulated trajectories, the policy is evaluated starting from each of the last T_i periods for $i = 1, \dots, H$ in the simulation. As a result, there are $N = \sum_{i=1}^{H} T_i$ states evaluated in the simulation.

To find the new parameters that approximate the value of the policy, the N simulated valuations of the consumption policy at the N states visited along the simulation trajectory are regressed on the basis functions of the polynomial at those states. That is, the sum of squares of the Bellman equation residuals visited along the simulation trajectory are minimised to give a new estimate of the approximate solution to the household's choice problem, $\hat{r}$.

As noted above, to implement the policy improvement step a dampened update of $\hat{r}$ is needed

$$r_{k+1} = (1 - \alpha_k)\, r_k + \alpha_k \hat{r} \tag{23}$$

where $\alpha_k \in (0, 1)$.[41] This update can then be inserted in eq. (21) to define the new consumption policy function for evaluation in the next iteration.

[39] To calculate continuation values, price factors are forecasted using eq. (13) and one-node Monte Carlo is used integrate out price shocks from eq. (12).

[40] At one extreme, as $\lambda \to 1$ very many periods are simulated and this algorithm closely approximates to exact policy iteration. At the other, $\lambda = 0$ the valuation of the policy is approximated by the sum of the flow utility at the current state and the discounted continuation value. I set $\lambda = 0.95$.

[41] The step size used can depend on the iteration k of the ADP algorithm. See Powell (2011) chapter 11 for a discussion of step size rules.

In implementation, partial updates using stochastic projected gradient descent methods are used to address numerical stability issues that arise from the high-dimensional, sparse nature of the household's choice problem.[42]

4 Econometrics

4.1 Estimation

The structural parameters, θ, are fitted using simulated methods of moments. Let $\hat{h}$ define the vector of moments observed in the data and $h(\theta, r)$ define the simulated moments at the current estimate of θ and r. Define the distance between the observed data moments and the simulated counterparts as $g(\theta, r) \coloneqq \hat{h} - h(\theta, r)$.[43]

To fit the structural parameters I solve the following optimisation problem

$$\begin{aligned} \theta^{\star}, r^{\star} &= \arg\min_{\theta, r} g(\theta, r)^{\top} \Sigma^{-1} g(\theta, r) \\ s.t. \quad & \Phi r = \Pi T(\Phi r) \end{aligned}$$

where Σ is a positive definite symmetric weighting matrix and Π is the projection with respect to the Euclidean norm, $\|\cdot\|$.

To implement the estimation I use an adaptive Markov chain Monte Carlo derivative free optimisation (Chernozhukov and Hong (2003); Łącki and Miasojedow (2015)). Briefly, this method uses multiple chains, each with different temperatures, proposal covariance scaling factors, and estimates for the dynamic demand model. The coldest chain is used to calculate the posterior density of the structural model's parameters.

As noted by Imai et al. (2009) and Norets (2009) solving large scale dynamic models at every parameter guess can be very costly. Their suggested remedy is to alternate between iterations of fitting the structural parameters and updates of the solution to dynamic program.[44] In line with this approach, I alternate between fitting the structural parameters, θ, and solving the dynamic demand model by doing a single iteration of λ-policy iteration algorithm described in Section 3.6. The details of this estimation are provided in Annex C.

[42] See Bertsekas (1999) and Parikh et al. (2014) for further details of stochastic projected gradient descent methods.

[43] The simulation of moments is as described in the Section 3.6 with the exception that length of the simulation used to calculate moments is fixed at 52 weeks. That is, starting from an initial inventory of zero, the first 12 weeks are excluded to reduce depends on initial inventories and the next 52 weeks of simulated purchases and consumption decisions are used to compute moments.

[44] Intuitively this avoids the costly step of fitting the solution to a dynamic program whose parameters may be far away from the true parameters. Moreover, they show that the alternating procedure converges to the same posterior distribution.

4.2 Identification

Like other dynamic demand models of storable goods, formal identification of the model is complex. As such, and in line with this existing literature, I provide an informal discussion of identification of model parameters.

In Section 2, data on inter-purchase durations, current and past prices, quantities purchased, and the sequences of SKUs chosen was used to demonstrate that price expectations, inventory holdings, and taste heterogeneity are important features of demand for laundry detergent in the UK. This data also identifies the parameters of the model.

The marginal utility of income is identified by standard arguments using variation in prices over markets. Shopping costs are identified by differences in the size of SKUs purchased across different types of household.

Inventory costs are identified by comparing inter-purchase durations of households. To illustrate, consider two households that always purchase detergent in one particular format. Over the same time period, they purchase the same number of washes. However, because one household has higher inventory costs than the other, they purchase smaller SKUs more frequently. As such, comparing inter-purchase duration for households with the same consumption rate identifies storage costs. To identify inventory costs differences between formats, this analysis can be conditioned on format purchased.

For a particular detergent j, the ratio ψ_j/Ψ_{jj} is an important determinant of the rate of consumption. Holding fixed ψ_j, Ψ_{jj} can be identified by comparing inter-purchase durations of households conditional on purchasing the same quantity of detergent j. To illustrate these ideas, consider two households who face the same prices and always buy the same number of washes. However, one household consistently purchases less frequently than the other. Therefore, over the same period the household that purchases more frequently will consume more washes. That is, it consumes detergent washes at a higher rate. Therefore, Ψ_{jj} can be identified by comparing inter-purchase durations of households conditional on purchasing the same quantity of detergent j. Conditioning this analysis on consecutive purchases for each detergent identifies all diagonal elements of Ψ.

In the above discussion, ψ_j is held fixed. This is because, in addition to impacting the rate of consumption, ψ_j directly impacts on the level of utility from consumption of detergent j. In turn, the linear utility weights are important parameters for matching market shares. As such, observed market shares will aid identification of ψ over and above consumption rates.

Interaction between quantities purchased, duration between purchases and identities of consecutive purchases of detergent help identify off diagonal terms in Ψ. This is because households that tend to use different detergents for different types of washing are likely to maintain more than one type of inventory. Since detergent is costly to store, all else equal, households with two sets of inventory will tend to purchase smaller SKUs of different inventories at close intervals. Therefore, by using the joint distribution of quantities purchased and inter-purchase duration of households whose consecutive purchases are of different detergents to those who purchase the same detergent helps identify off-diagonal terms in Ψ.

The scaling parameters on the random coefficients are identified by re-purchase probabilities.

5 Empirical Results

The model is estimated in two steps. First, I estimate the household's price forecasting model described in Section 5.1. Next, I estimate the dynamic demand model. The results of the dynamic demand model are presented in Section 5.2. In Section 5.3 I conduct some policy experiments using the model and use them to show how temporary price cuts can lead to purchase acceleration.

5.1 Price Model

Purchases between 2009 and 2011 which are recorded by Kantar Worldpanel are used to estimate the price forecasting model. The underlying assumption is that all households in the sample observe the same set of prices. This is an appropriate assumption for the UK, where large supermarkets must charge the same price for a given SKU throughout the country following a ruling by the UK Competition Commission in 2000.[45]

Because the data is only recorded when households make purchases, the price series for each SKU is only partially observed. While, the interactive fixed effects model can be used to impute missing prices, the panel is too unbalanced for this to work well in practice. Instead SKUs of a similar size are grouped together, because they are likely to exhibit similar promotional activity.[46] For each of the 37 types of detergent I assign SKUs to one of four groups: less that 18 washes, between 18 and 24 washes, 25 to 40 washes and more than 40 washes. This results in 112 groups of SKUs.

The interactive fixed effects model is estimated for up to three factors using non-linear least squares. In each case I estimate a vector autoregression with up to four lags.

To choose the optimal number of factors I use the information criteria proposed by Bai and Ng (2002). These are shown in the top panel of 1. Two of of the information criteria are minimised by a price model with two factors, the other with three factors.

The bottom panel reports the results of the Schwarz-Bayes information criterion (SBIC) when up to four lags are included in the vector-autoregression for the factors from each of the interactive fixed effect models. When there is one factor, the SBIC is minimised by including two lags. Otherwise, the SBIC is minimised by choosing one lag.

Taken together, these criteria suggest that either a two or three factor model with one lag would could be used. Since there is an additional premium to increasing the state

[45]Some price variation is permitted for 'small' supermarkets that are similar in size to convenience stores (less than 280 sq m).

[46]For example, very large SKUs are offered at quantity discounts and are therefore not promoted very regularly.

Table 1: Price Model: interactive fixed effects model

Number of Factors	R=1	R=2	R=3	$R^{\star}$
ln(Obj. Func.)+R×Pen. Fn.				
Penalty 1: $\frac{N+T}{NT}log(\frac{NT}{N+T})$	-2.064	-2.109	-2.102	2
Penalty 2: $\frac{N+T}{NT}\log(C)$	-2.052	-2.084	-2.065	2
Penalty 3: $\log(C)/C$	-2.094	-2.170	-2.193	3
Number of Lags, SBIC				
τ=1	-8.241	-16.620	-25.481	
τ=2	-8.242	-16.574	-25.321	
τ=3	-8.227	-16.492	-25.089	
τ=4	-8.190	-16.364	-24.848	
$\tau^{\star}$	2	1	1	

Note: $C = \min\{N, T\}$

space I opt for the 2 factor model with 1 lag.[47] The results of this interactive fixed effect model and the $VAR(1)$ applied to the factors are shown in Tables 2 and 3, respectively.

When translated back into a price per wash, 90 percent of the fitted average price per wash for the 112 groups are within 2% of the average price per wash in the data. Therefore, the model fits average prices well over brands, formats and sizes.

5.2 Dynamic Demand Model

The model is estimated using household purchase diary data described in Section 2. For an individual household, the parameter space is $\theta_i = [\psi_i, \Psi_i, \psi_{i,0}, \rho_0, \gamma_i, \Gamma_i, \kappa_{i,1}, \kappa_{i,2}]$ with $J^2 + J + 3$ parameters. With 37 detergents, there are approximately 1,400 parameters. To ensure that the resulting model is estimable, a more parsimonious parameterisation of the model is needed.

5.2.1 Parameterisation

Households' flow utility is slightly modified for estimation. There are two main alterations to eq (4). First the marginal utility of income is included in the parameters of the purchase and inventory costs. Second, since income is constant across purchase alternatives it is omitted from the model. The resulting flow utility function is

$$U(C_{it}) - \psi_{i0} P_{mt} - PC_{it} - IC_{it} + \frac{1}{\sigma_\varepsilon}\varepsilon_{imt} \tag{24}$$

[47] The results of coefficients of all three interactive fixed effect models are shown in Annex B.

Table 2: Factor Model: Interactive fixed effect model with 2 factors

		Estimate	Std. Error
Formats (excl. Caps):			
	Powder	-0.449	(0.119)
	Liquid	-0.709	(0.148)
	Tablets	-0.574	(0.147)
	Gel	-0.589	(0.119)
Brands (excl. Brand A)			
	Brand B	0.068	(0.110)
	Brand C	-0.539	(0.106)
	Brand D	0.277	(0.106)
	Brand E	-0.688	(0.170)
	Brand F	-1.032	(0.175)
	PL	-1.069	(0.112)
	Others	0.377	(0.243)
Other Chars (excl. Single SKU)			
	Multi-Pack	0.007	(0.118)
	Washes	-0.383	(1.708)
	Washes^2	-1.106	(2.439)
	Washes^3	0.808	(1.267)
Constant		0.689	(0.196)
R^2: Overall		0.710	
R^2: Within		0.132	
Num. Obs.		14,927	

Table 3: Factor Model: VAR(1) with 2 factors

	$F_{1,t}$		$F_{2,t}$	
	Estimate	Std. Err	Estimate	Std Error
$F_{1,t-1}$	0.795	(0.067)	0.192	(0.163)
$F_{2,t-1}$	0.054	(0.025)	0.884	(0.061)
Constant	0.017	(0.006)	-0.018	(0.014)

Turning to the parameters themselves, for household i the linear utility weight for each detergent j is

$$\psi_{ij} = h_\psi\left(\bar{\psi}_j + \omega_{ij}\right) \tag{25}$$

where $h_\psi(\cdot)$ is scaled logistic transformation applied to ensure that the weights are strictly positive, $\bar{\psi}_j$ is a detergent specific component common to all households, and ω_{ij} is a household specific component.[48]

The household specific component is a draw from the unobserved distribution of persistent taste heterogeneity. Ideally, the taste draws for each detergent would be included in the state vector to capture the effect that these preferences have on the expected value of holding different detergent. However, with 37 different types of detergent, doing so would reintroduce to the curse of dimensionality.

To resolve this issue I use a factor structure to provide a low-dimensional approximation to the high-dimensional distribution of persistent taste heterogeneity. Specifically, I assume a household's preference for each detergent is a function of two standard normal shocks, $\eta_{k,i} \sim N(0,1)$, $k = 1,2$. To map the two shocks into J detergent specific taste parameters for each household, the shocks are multiplied by brand and format specific weights $\sigma_{k,b}, \sigma_{k,f} \geq 0$ for $k = 1,2$. The household specific component is,

$$\omega_{ij} = \sum_{k=1}^{2} \sum_{b \in \mathcal{J}_B} \sum_{f \in \mathcal{J}_F} \left(\sigma_{k,b} + \sigma_{k,f}\right) \eta_{k,i} \tag{26}$$

where $\mathcal{J}_B$ and $\mathcal{J}_F$ are the sets of brands and formats, respectively.

The (j,k)-th entry for quadratic utility weights for all $j,k = 1,\ldots,J$ are

$$\ln \Psi_{ijk} = \frac{1}{2}\ln\psi_{ij} + \frac{1}{2}\ln\psi_{ik} + \ln\upsilon_{jk} - \ln\left(\alpha_1 + \alpha_2 Z_i\right) \tag{27}$$

where Z_i is the number of equivalent adults in household i. Further, $\upsilon_{jj} = 1$ and $\upsilon_{jk} = \upsilon \in (0,1)$ for all $j \neq k$. The final term allows the consumption rate to vary with household size.

The price coefficient is allowed to vary across households and is a function of income per equivalent adult

$$\ln\psi_{i0} = \mu_0 - \sigma_0 \frac{\tilde{Y}_i}{Z_i} \tag{28}$$

where $\frac{\tilde{Y}_i}{Z_i}$ is a standardised income per equivalent adult in household i, $\sigma_0 > 0$. In the estimation I impose the location normalisation that the household with the average income per adult is 1. That is, I fix $\mu_0 = 0$.

Noting the similar role of both inventory cost parameters, I estimate Γ and fix $\gamma = 0$.

[48] Specifically, $h_\psi(x) = 10\left(1 + \exp\left(-\frac{x}{3}\right)\right)^{-1}$.

5.2.2 Results

In this section I present the results of the model.[49] The results of the estimation are shown in Table 4 and are split into three panels. The top panel reports the value of the linear utility weight evaluated at $\eta_1 = \eta_2 = 0$, the middle panel shows the brand and format weights on both taste shocks, and the lower panel shows all other parameters.

The top two panels contain the parameter estimates for the linear utility weights, which show that the common detergent parameters and scaling factor on taste vary across brands and formats. However, in a non-linear model the parameters are difficult to directly interpret. So while utility parameters are observed to vary across brands and formats, the effect they have in the model is difficult to establish.

Instead, to evaluate the fit, it is instructive to compare the moments predicted by the model to those observed in the data. To calculate moments implied by the model I conduct a simulation at the estimated parameter values as described in Section 3.6. There are four main elements of the model: the brand shares, format shares, re-purchase probabilities and the distribution of inter-purchase durations. Each is discussed in turn.

Figure 8 plots the brand purchase shares in the simulation of the model against those in the data. The shares are split into the four quartiles of the distribution of income per equivalent adult. The top (bottom) row of plots shows the brand shares of the upper (lower) half of the income per equivalent adult distribution. The marker indicates the identity of the brand purchased and the red line is the 45 degree line.

Figure 8 shows that the simulated brand shares captures some important features of the observed data. First, as in the observed data, branded SKUs are especially popular with the highest income households. Second, the retailer's private label is strongly preferred by the least well off households. However, the simulated share of PL detergent is a overstated for the poorest households, and slightly understated for all other income quartiles. Finally, with the exception of Brand A, the other brand shares fit the observed data well, especially for Brand E - the leading brand.

Figure 9 plots the actual versus the simulated share by the detergent format purchased. Again, the top (bottom) row shows the simulated versus the observed shares of different formats for the richest (poorest) households. The marker indicates which format is purchased and the 45 degree line is shown in red.

The figures shows that the simulated purchase shares for formats closely align with the data. Specifically, as in the observed data, powder is the most preferred format in all income quartiles. The model also captures the fact that the cheapest format, liquid, has a higher share for poorer households. Further, the simulated shares also reflect the fact that the most expensive format, capsules, is most popular with richer households. The simulated shares of gel coincide with those in the data, while the simulated shares of tablets is slightly understated by the model.

Next, I explore a prominent feature of the observed data discussed in Section 2.4 - the re-purchase probabilities. Figure 10 plots the re-purchase probabilities in the data against those from the simulation and it contains two panels.

[49] I use the last 100 iterations of the estimation algorithm to calculate parameter estimates and confidence intervals.

Table 4: Model parameters

Linear Utility: $\psi_{ij} : \eta_1 = \eta_2 = 0$

Brands	**Formats** Capsules	Gel	Liquid	Powder	Tablets
Brand A	0.514	2.058	0.343	8.136	3.345
	[0.150,0.787]	[1.549,3.052]	[0.096,0.549]	[7.141,9.591]	[0.078,6.545]
Brand B	1.567	1.756	4.383	0.747	0.572
	[0.641,2.330]	[1.066,4.384]	[3.346,6.078]	[0.044,1.434]	[0.310,0.987]
Brand C	1.168		0.469	0.334	0.756
	[0.257,4.664]		[0.285,0.712]	[0.153,0.557]	[0.534,0.998]
Brand D	1.037	0.397	2.555	1.569	0.509
	[0.017,2.616]	[0.167,1.317]	[0.608,4.031]	[1.096,3.062]	[0.083,2.662]
Brand E	2.465		7.418	1.809	0.364
	[0.977,3.473]		[4.898,9.567]	[0.554,3.437]	[0.016,2.519]
Brand F	1.355		1.402	1.529	1.077
	[0.708,2.811]		[1.041,1.673]	[1.231,2.069]	[0.155,1.684]
PL	0.871	2.391	0.656	2.103	3.867
	[0.306,2.287]	[1.28,2.985]	[0.05,1.495]	[0.018,4.365]	[1.5,8.495]
Other	3.675	1.255	2.956	2.292	2.236
	[0.090,6.805]	[0.092,2.338]	[0.189,9.88]	[0.498,3.639]	[1.015,3.613]

	σ_1	σ_2		σ_1	σ_2
Brands:			**Formats**:		
Brand A	0.809	1.177	Capsules	1.157	0.985
	[0.224,3.435]	[0.779,1.661]		[0.632,3.009]	[0.107,1.953]
Brand B	0.396	1.708	Gel	1.184	1.159
	[0.150,1.201]	[0.638,3.786]		[0.754,1.688]	[0.764,1.475]
Brand C	1.656	2.676	Liquid	1.157	2.363
	[1.375,1.89]	[0.956,3.312]		[0.791,1.873]	[0.924,3.074]
Brand D	0.096	0.047	Powder	0.693	3.576
	[0.042,0.388]	[0.017,0.081]		[0.556,0.977]	[1.944,3.928]
Brand E	1.297	2.000	Tablets	0.405	0.976
	[0.036,2.181]	[0.920,3.840]		[0.002,2.527]	[0.503,1.270]
Brand F	0.580	1.494			
	[0.122,1.105]	[0.879,3.261]			
PL	0.456	2.067			
	[0.351,0.664]	[1.206,3.385]			
Other	0.234	0.969			
	[0.084,0.341]	[0.132,3.868]			

Other Parameters

Quad. Utility: Ψ_i		ψ_{i0} **& Shopping Costs**		**Inventory Costs**	
υ	0.997	μ_0	0	γ	0
	[0.992,1]		-		-
α_1	1.702	σ_0	3.031	Γ	0.025
	[0.808,1.930]		[0.300,3.875]		[0.003,0.041]
α_2	0.073	ρ_0	7.06	κ_1	0.790
	[0.030,0.120]		[5.577,8.162]		[0.132,3.45]
		Utility Shocks		κ_2	10.42
		σ_ε	0.073		[6.404,17.491]
			[0.043,0.154]		

Note: 90% confidence intervals in square brackets

Figure 8: Brand shares by quartiles of income per equivalent adult: simulated vs. data

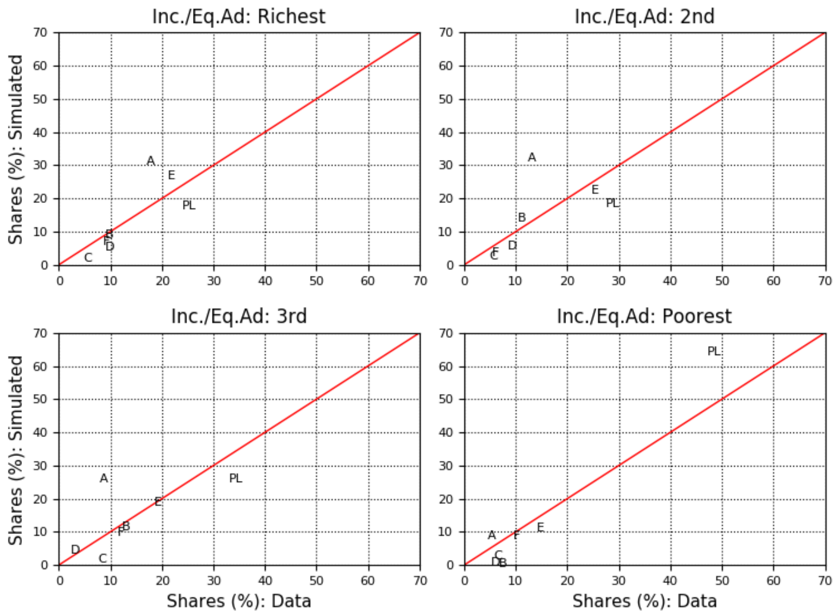


Figure 9: Format shares by quartiles of income per equivalent adult: simulated vs. data

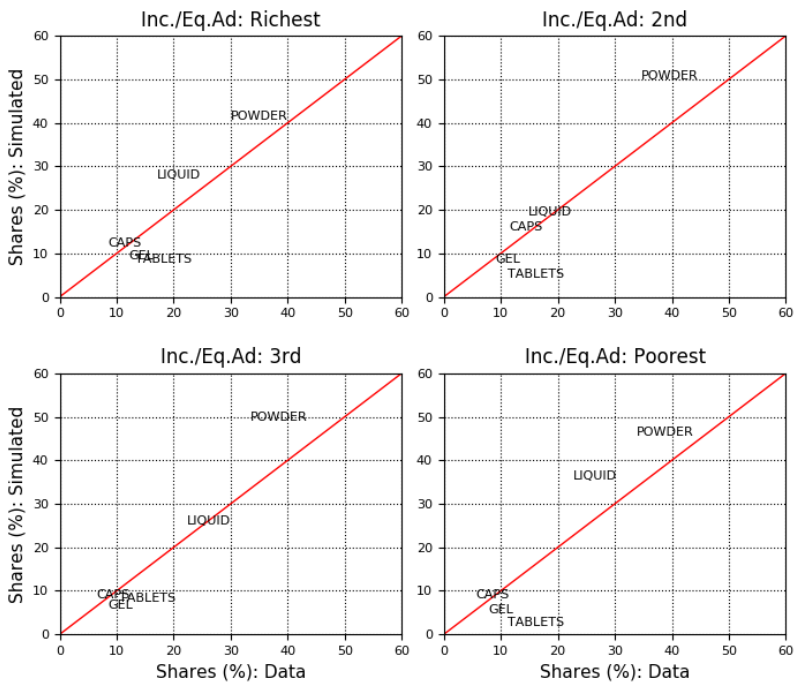
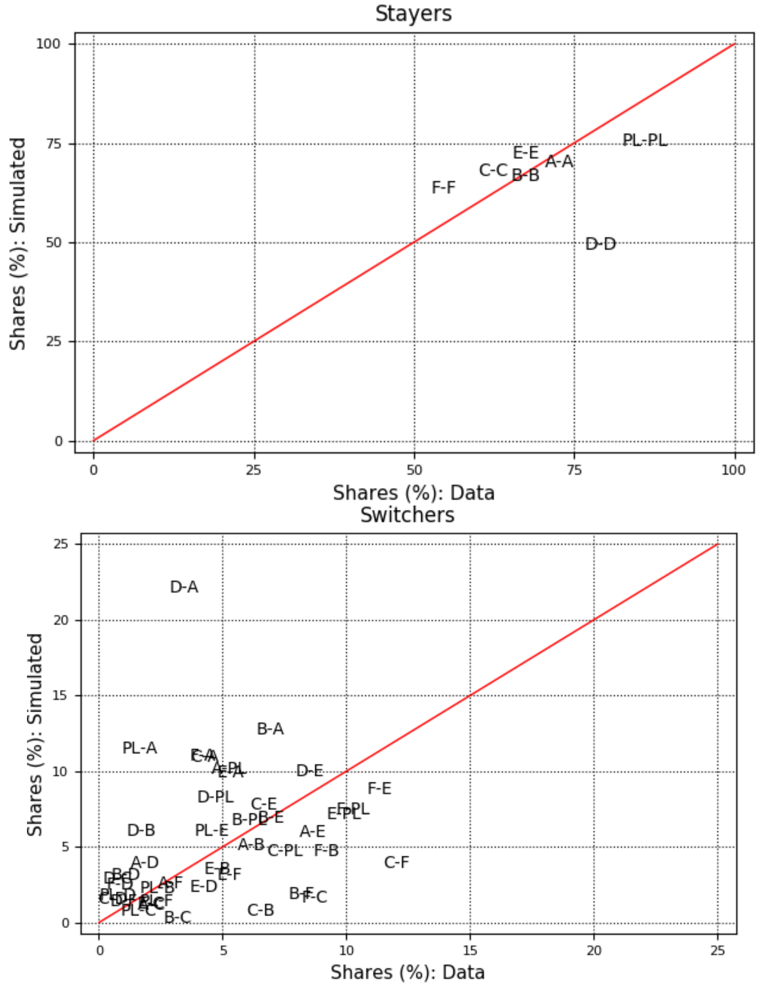


Figure 10: Brand re-purchase probabilities: simulated vs. data



The top panel plots the percentage of households that buy the same brand as their previous purchase. I refer to these brand loyal customers as 'stayers'. In general, the model successfully predicts the high re-purchase rates observed in the data - only Brand D's repurchase rates are too low.

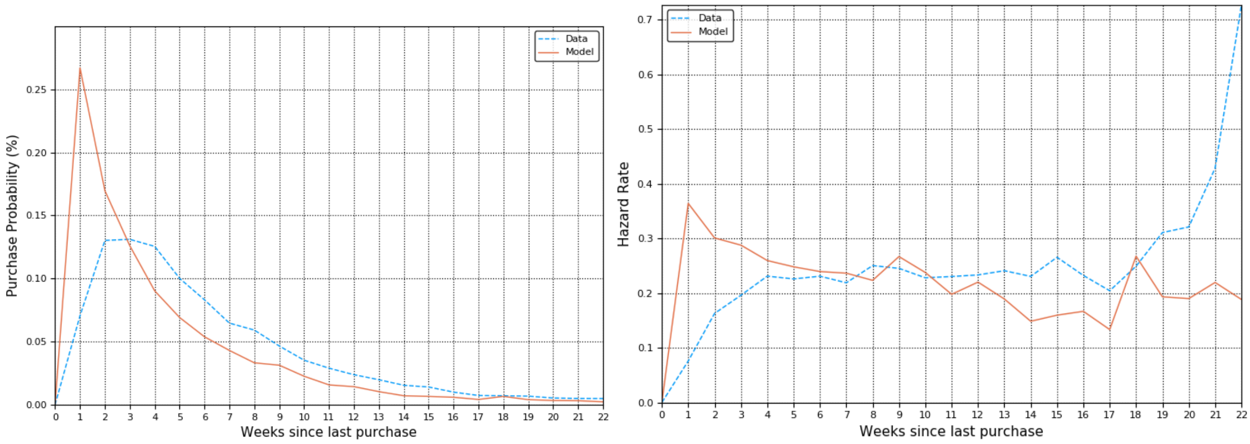
The bottom panel presents the re-purchase rates for switchers. In general, the fit is once again good. With only a handful of exceptions, the simulated the re-purchase probabilities are clustered around the 45 degree line. Qualitatively similar results hold for formats.[50]

This analysis suggests that although low-dimensional, the approximate factor structure used to capture persistent unobserved taste heterogeneity across brands and formats is effective at reproducing observed purchase dynamics.

Figure 11 shows the purchase probabilities and hazard rates for inter-purchase durations in the model and in the data. The blue dashed lines correspond to the data and the red solid lines are the simulated output from the model.

[50]See Annex F.

Figure 11: Fit of purchase probability and hazard rate of inter-purchase duration



The simulated inter-purchase duration broadly captures the key features of the inter-purchase duration distribution observed in the data. Namely, most purchases occur in the first few weeks after the last purchase and the remainder are spread over the next 15 to 20 weeks.

However, there are some notable differences between the simulated and observed distributions. Namely, the simulated number of purchases in the first two weeks following the previous purchase is considerably higher than that implied by the data - illustrating that the simulated hazard rate is too high in the first few weeks following a purchase. As a result, the percentage of purchases is overstated in weeks 1 and 2, and understated in the medium to long-run.

Inspection of simulation output indicates this occurs because households tend to buy several small SKUs of the same detergent in sequential weeks, rather than one large SKU. That is, households make too many purchases, too close together.

One possible remedy is, as suggested in Section 3.2.2, to accrue inventory costs for the number of SKUs, rather than number of detergents. As noted earlier, with the addition of an inventory consumption rule (ie. first-in-first-out) this can be relatively easily incorporated into the model.

5.3 Policy Experiments

In this section I present two types of policy experiments. First, I examine the impact of transitory price changes. This illustrates how promotional pricing can lead to purchase acceleration. Second, I calculate medium to long-run price elasticities for the most popular SKUs by simulating the demand response to a permanent price increase.

5.3.1 Demand response to transitory price changes

I report the results of two experiments in which there is an unexpected, short-lived price cut for some of most popular branded SKUs. Using these experiments I show how short-run price changes can impact on the timing and identity of detergent purchased.

To conduct each experiment, I first generate 12,400 households that are representative of the observed sample of households.[51] Each household appears in the experiment for 52 weeks. Observed SKU prices are used in the simulation, as are the price factors estimated using eq. (12).

In the 27th week after the simulated household first arrives in the sample, I reduce the observed prices of the selection of SKUs by 10 percent. Because the week in which the price is cut is relative to the point at which each household enters the sample, the effect of the price cut is integrated out over the range of observed prices.

The simulation procedure mirrors the one used to calculate the objective function. The impact of the price cut is evaluated by tracing out the impulse response function of the SKUs purchased during and after week 27.

In the price forecasting model similar sized SKUs are grouped together and are governed by the same price dynamics. I conduct this experiment for Brand A powder SKUs and Brand E liquid SKUs containing 25 to 40 washes.

Figure 12 shows the demand response of a 1 week price cut of 10 percent for large Brand A powder SKUs. It contains three panels. The top panel shows the change in the number of washes purchased for the promoted SKUs. The same plot is repeated for other Brand A powder SKUs and all other detergents in the middle and bottom panels, respectively.

The top panel in the figure shows that households in the simulation bought around 130 more washes of Brand A powder in large SKUs in response to the price cut. This initial demand increase occurs without immediately reducing purchases of other detergents. However, in the periods following the price cut, the future purchases of other detergents are reduced. This suggests that purchases were accelerated and switched to Brand A powder in response to the unexpected price cut.

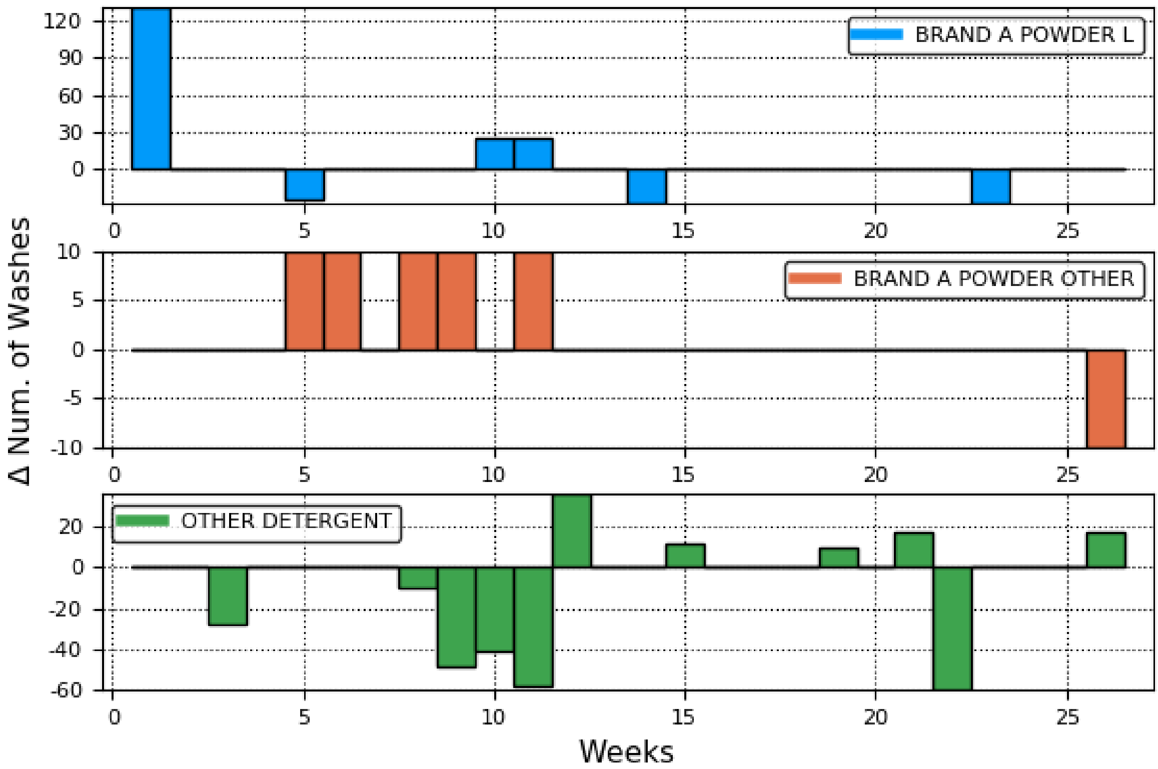
There is also a smaller increase in the demand for Brand A powder SKUs in weeks 5 to 10 of the price experiment. This demonstrates that once households switch to Brand A powder they are more likely to buy other SKUs containing the same detergent.

Figure 13 shows the change in the washes purchased for each detergent when there is a temporary one week price cut of 10 percent for all large SKUs of Brand E liquid. It contains three panels that plot the impulse response functions of number of washes purchased in large Brand E liquid SKUs (top), other Brand E liquid SKUs (middle), and other detergent SKUs (bottom).

The top panel in the figure shows that households in the simulation bought 56 more washes of Brand E liquid in large SKUs in the week of the price cut. Again, the immediate demand response to the price cut occurs without reducing contemporaneous purchases of other detergents. However, in the periods following the price cut, the purchases of other detergents are reduced. This inter-temporal substitution suggests that purchases were accelerated and switched to Brand E liquid in response to the unexpected price cut.

[51] Specifically, I make 20 versions of each household each with different draws for persistent taste heterogeneity and SKU-specific utility shocks.

Figure 12: Unexpected 10% price cut for 1 week: Brand A Powder SKUs (25-40W)



There is also a increase in demand for Brand E liquid SKUs in weeks 6 and 8 of the price experiment. As in the first price experiment, this demonstrates households are more likely to buy other SKUs containing the same detergent if they already have them in stock. Through this mechanism, promotional pricing has the potential for long run increases in volumes sold for the promoted good.

5.3.2 Demand response to permanent price changes

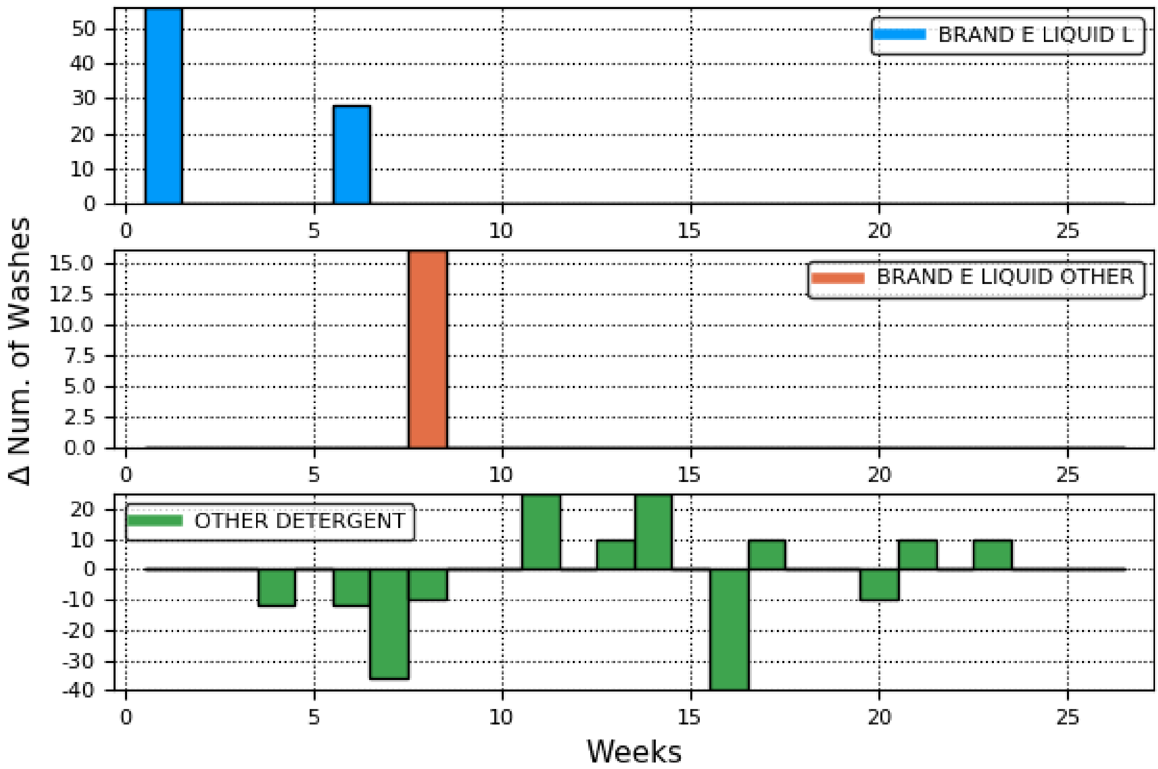
Next, I simulate the impact of a permanent one percent price rise for 12 SKUs. These 12 SKUs represent a selection of popular SKUs with different formats, brands and sizes. I conduct the price experiment separately for each SKU.

Like the short-run price experiment, I begin by simulating 12,400 households. Again, the households are representative of the households in the sample of Kantar Worldpanel data used in estimation.

Observed prices are used in the simulation. However, price expectations are adjusted in line with the permanent price rise - which implies a revision to households' forecasting model. To adjust the price model for the permanent price change I multiply the cross-sectional component of the price forecasting model for SKU m by k_m that solves

$$1.01\bar{p}_m = \sum_{t=1}^{T} \int z^{-1} \left(\lambda_m^\top F_t + k_m X_m^\top \alpha + \epsilon_{imt} \right) dG_\epsilon \tag{29}$$

Figure 13: 10% price drop over 1 week: Brand E Liquid SKUs (25-40W)



and $\bar{p}_m$ is the observed average price per wash for SKU m in the data. In each period, the observed prices for SKU m are shifted up by 1 percent of its average price over the sample. This leaves the time series variation of prices unchanged.

Before simulation, the model is solved assuming households use the adjusted forecasting model and corresponding prices. With the new price forecasting model, counterfactual prices and model solution, the simulation is conducted as it was in the short-run policy experiment.

The own-price elasticities are reported in Table 5. The first column shows the own-price elasticities for all households. It shows that the own-price elasticities are greater than -1, and detergent would appear to be inelastically demanded.

To understand why this is the case, I split the sample by income per equivalent adult. The last two columns in Table 6 report the own-price elasticities for the poorest and richest households separately.

For the households with below median income per equivalent adult, the own price elasticity is greater than 1 in absolute value for 10 out of 12 SKUs. The larger branded SKUs tend to be more elastically demanded by the poorest households: Brand A gel 28W, Brand B capsules 20W, Brand A powder 42W, and Brand E liquid 28W all have elasticities less than -3. Perhaps reflecting the ability of households with lower inventory costs to better time their purchases with sales of their preferred detergent before running out of stock.

Table 5: Own-price elasticities

SKUs	Own-price elasticities All Households	Poorest	Richest
Brand A Gel 28W	-0.336	-3.211	0.308
Brand B Caps 20W	-0.394	-3.030	0.207
Brand E Liquid 16W	-0.410	-1.951	0.139
Brand E Liquid 18W	0.066	-0.937	0.414
Brand E Liquid 28W	-0.828	-4.933	0.231
Brand F Liquid 18W	-0.712	-1.280	-0.377
Brand A Powder 10W	0.140	-0.264	0.330
Brand A Powder 25W	-0.073	-1.477	0.360
Brand A Powder 42W	-1.634	-3.097	-1.322
PL Powder 10W	-0.451	-1.056	0.000
PL Powder 30W	-0.528	-1.241	0.200
PL Tablets 12W	-0.924	-2.345	-0.121

Smaller branded SKUs tend to be more inelastically demanded - perhaps reflecting the increased market power manufacturers have over households with relatively high inventory costs. Finally, the retailer's private label SKUs are also quite inelastically demanded - this is consistent with the highly loyal customer base.

Counterintuitively, for the wealthiest households, there is often a small positive demand response to the price rise. This marginal increase in purchases for a small number of households reflects the fact that the cost of running out of stock during a non-promotional period has now increased from an already high base price. Absent the price increase, some households decide not to purchase in some of the promotional periods. They prefer to consume from existing stock and wait for the next promotion - even though they may incur a cost for running out of stock during a non-promotional period (i.e. either through a higher purchase price, or through reduced consumption). Following the price increase, for a small number of households it is now preferable to purchase SKUs more frequently when they are on sale and keep stocks at a higher level. This enables them to avoid with certainty the additional cost they would otherwise incur if they run out of stock before the next promotion occurs.

Table 6 shows the cross-price elasticities implied by the model. The first of these shows the substitution to different size SKUs containing the same detergent. The remaining two columns show substitution to other detergents and the option not to purchase.

Taken together these cross-price elasticities show that different size SKUs containing the same detergent are particularly close substitutes. This reinforces the importance of household preferences for specific detergents in understanding consumer dynamic purchasing behaviour.

Table 6: Cross-price elasticities

SKUs	Cross-price elasticities Same det., diff size	Other detergent	No Purchase
Brand A Gel 28W	0.394	-0.016	0.002
Brand B Caps 20W	0.000	0.003	0.001
Brand E Liquid 16W	0.016	-0.008	0.003
Brand E Liquid 18W	0.135	-0.006	0.001
Brand E Liquid 28W	0.052	-0.011	0.003
Brand F Liquid 18W	0.193	0.006	0.000
Brand A Powder 10W	0.000	0.026	-0.006
Brand A Powder 25W	0.050	-0.002	-0.001
Brand A Powder 42W	-0.008	0.048	-0.003
PL Powder 10W	-0.072	0.015	0.003
PL Powder 30W	0.052	-0.008	0.003
PL Tablets 12W	-0.602	0.012	0.000

5.3.3 Summary

These price experiments highlight the role of purchase acceleration from promotional prices, suggest that inventory costs confer market power on manufacturers, and highlight the importance of unobserved heterogeneity in preferences in understanding consumer dynamics.

They also highlight that the model does not produce enough price sensitivity for wealthy households. This is most likely related to the lack of flexibility in the specification of the marginal utility of income. This aspect of the model is being investigated further in ongoing research.

6 Conclusion

This paper develops a dynamic discrete-continuous demand model for storable goods - a class of fast moving consumer goods that account for a large fraction of grocery expenditures. It is applied to the UK laundry detergent industry using household level purchase data.

To estimate and solve the dynamic demand model, I use techniques from: approximate dynamic programming, large scale dynamic programming in economics, machine learning, and statistical computing. The benefits of this approach are three-fold.

First, the dynamic demand model is compatible with high-dimensional choice sets. In turn, making dynamic demand estimation possible for storable good industries with many sizes - the UK laundry detergent industry is an example.

Second, the model can combine the most desirable features of existing models. In particular, it allows for persistent taste heterogeneity to interact with product varieties in flow utility and continuation values. Furthermore, utility from product differentiation accrues at the point of consumption, not purchase. Together these features enable the model to capture rich inter- and intra-temporal substitution patterns.

Finally, these dimension reduction techniques do not hinge on idiosyncratic features of the industry being studied, nor do they impose restrictive assumptions on purchase decisions and/or consumption. As a result, this dynamic demand can be applied to any storable good industry with only minor modifications.

This model is likely to be of both policy and commercial interest. In a policy setting, understanding how consumers react to price dynamics may be important for effective design of taxation policy. In addition, consistent estimation of short and medium to long run elasticities is a key input into antitrust analysis of mergers, assessment of cartel damages, etc. Finally, these storable demand models can be used to construct new cost-of-living indices to reflect differences in the prices recorded in baskets of goods and purchase prices (Osborne (2017)).

From a commercial perspective, this structural dynamic demand model can be applied to consumer level purchase data using ever increasing computational resources. The resulting demand model enables firms to better understand demand dynamics - a key input into the optimisation of promotional price strategies and demand forecasting. Moreover, it provides a new way to explore counterfactual market outcomes when new products are introduced or old products are withdrawn.

References

Arcidiacono, P., Bayer, P., Bugni, F. A., and James, J. (2012). Approximating high-dimensional dynamic models: Sieve value function iteration. Working Paper 17890, National Bureau of Economic Research.

Arellano, C., Maliar, S., Maliar, L., and Tsyrennikov, V. (2014). Envelope condition method with an application to default risk models.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*, 77(4):1229–1279.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, 3(2):89–163.

Baragatti, M., Grimaud, A., and Pommeret, D. (2013). Likelihood-free parallel tempering. *Statistics and Computing*, 23(4):535–549.

Benitez-Silva, H., Rust, J., Hitsch, G., Pauletto, G., and Hall, G. (2000). A Comparison Of Discrete And Parametric Methods For Continuous-State Dynamic Programming Problems. Computing in Economics and Finance 2000 24, Society for Computational Economics.

Bertsekas, D. (2011a). *Approximate Dynamic Programming*, volume II. 3rd edition.

Bertsekas, D. (2015). Lambda policy iteration: A review and new implementation. *Reinforcement Learning and Approximate Dynamic Programming for Feed Control.*

Bertsekas, D. and Ioffe, S. (1996). Temporal differences-based policy iteration and applications in neuro-dynamic programming. Technical report, Lab. for Info. and Decision Systems Report LIDS-P-2349, MIT.

Bertsekas, D. P. (1999). *Nonlinear programming*. Athena scientific Belmont.

Bertsekas, D. P. (2011b). Dynamic programming and optimal control 3rd edition, volume ii. *Belmont, MA: Athena Scientific.*

Boizot, C., Robin, J.-M., and Visser, M. (2001). The demand for food products: An analysis of interpurchase times and purchased quantities. *The Economic Journal*, 111(470):391–419.

Brumm, J. and Scheidegger, S. (2015). Using adaptive sparse grids to solve high-dimensional dynamic models. *Available at SSRN 2349281.*

Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Dubin, J. A. and McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, pages 345–362.

Erdem, T., Imai, S., and Keane, M. (2003). Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics*, 1:5–64.

Farias, V., Saure, D., and Weintraub, G. Y. (2012). An approximate dynamic programming approach to solving dynamic oligopoly models. *The RAND Journal of Economics*, 43(2):253–282.

Fowlie, M., Reguant, M., and Ryan, S. P. (2016). Market-based emissions regulation and industry dynamics. *Journal of Political Economy*, 124(1):249–302.

Hendel, I. and Nevo, A. (2006a). Measuring the implications of sales and consumer inventory behaviour. *Econometrica*, 74(6):1637–1673.

Hendel, I. and Nevo, A. (2006b). Sales and consumer inventory. *RAND Journal of Economics*, 37(3):543–561.

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

Judd, K., Maliar, L., Maliar, S., and Valero (2014). Smolyak method for solving dynamic economic models - lagrange interpolation, anisotrpohic grid and adaptive domain. *Journal of Economic Dynamic and Control.*

Judd, K. L., Maliar, L., and Maliar, S. (2011). Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models. *Quantitative Economics*, 2(2):173–210.

Łącki, M. K. and Miasojedow, B. (2015). State-dependent swap strategies and automatic reduction of number of temperatures in adaptive parallel tempering algorithm. *Statistics and Computing*, pages 1–14.

Maliar and Maliar (2014). Numerical methods for large scale dynamic models. *Handbook of Computational Economics.*

Maliar, L. and Maliar, S. (2013). Envelope condition method versus endogenous grid method for solving dynamic programming problems. *Working Paper.*

Malin, B. A., Krueger, D., and Kubler, F. (2011). Solving the multi-country real business cycle model using a smolyak-collocation method. *Journal of Economic Dynamics and Control*, 35:229 – 239.

Melnikov, O. (2000). Demand for differentiated durable products: The case of the us computer printer market. *Manuscript. Department of Economics, Yale University.*

Melnikov, O. (2013). Deman for differentiated durable products: The case of the u.s. computer printer market. *Economic Inquiry*, 51(2):1277–1298.

Nevo, A. and Hendel, I. (2012). Intertemporal price discrimination in storable goods markets. *Working Paper*.

Norets, A. (2009). Inference in dynamic discrete choice models with serially correlated unobserved state variables. *Econometrica*, 77(5):1665–1682.

Osborne, M. (2013). Esimation of cost-of-living-indices for a storable good using a dynamic demand model. *Working Paper*.

Osborne, M. (2017). Approximating the cost-of-living index for a storable good. *Rotman School of Management Working Paper*.

Parikh, N., Boyd, S., et al. (2014). Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239.

Powell, W. B. (2011). *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, volume 842. John Wiley & Sons.

Stock, J. and Watson, M. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, pages 1167–1179.

Sun, B. (2005). Promotion effect on endogenous consumption. *Marketing science*, 24(3):430–443.

Sweeting, A. (2013). Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry. *Econometrica*, 81(5):1763–1803.

Wang, E. Y. (2012). Estimating the distributional impact of taxes on storable goods: A dynamic demand model with random coefficients. *Working Paper, University of Massachussets*.

Wang, E. Y. (2013). The effect of more accurate cost-of-living indexes on welfare recipients. *Working Paper*.

Wang, E. Y. (2015). The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity. *The RAND Journal of Economics*, 46(2):409–441.

Annex A: Data

Additional filters are added to the sample of households to make them suitable for use in estimation of a dynamic demand model. To ensure households purchase records are likely to be informative for an analysis demand dynamics, they are required to make at least 10 purchases with at most one purchase per week. To guard against including households who temporarily drop out of the sample, the maximum gap between any two purchases is 24 weeks.

In addition, to only include households whose purchases are for personal consumption, any households that purchases more than 100 washes or buy more than 2 packs of detergent in a single shopping trip are omitted. In total, the filtered sample contains 620 households.

Annex B: Equivalence of the two stage problem

Let the pre-purchase state space be denoted by $x_{it} = \left[I_{it}, \tilde{J}_{it}, P_t, Y_{it}\right]$ where $\tilde{J}_{it} = \sum_{j=1}^{J} 1\left[I_{ijt} > 0\right]$ and the state space after purchases of SKU m be $s_{it} = \left[\bar{I}_{it}, \bar{J}_{it}, P_t, \bar{Y}_{it}\right]$. In recursive form, the household's choice problem is

$$\begin{aligned} V\left(x_{it}, \varepsilon_{it}\right) &= \max_{d_{it}} \max_{0 \le C_{it} \le I_{it}+Q_{it}} U\left(C_{it}\right) + \psi_{i,0}\left(Y_{it} - P_{it} - PC_{it} - IC_{it}\right) \\ &\quad + \sum_{m=0}^{M} d_{imt}\varepsilon_{imt} + \delta \int V\left(x_{it+1}, \varepsilon_{it+1}\right) dG_{\varepsilon} dG_{P_{t+1}|P_{t,\tau}} \end{aligned} \tag{30}$$

where $P_{t+1} \sim G_{P_{t+1}|P_{t,\tau}}$, $d_{it} = \left[d_{i0t}, \ldots, d_{iMt}\right]^{\top}$, $\varepsilon_{it} = \left[\varepsilon_{i0t}, \ldots, \varepsilon_{iMt}\right]^{\top}$ and $\sigma_{\varepsilon} = 1$ for simplicity.

Integrating the value function over SKU specific shocks, ε_{imt},

$$\bar{V}\left(x_{it}\right) := \int V\left(x_{it}, \varepsilon_{it}\right) dG_{\varepsilon} \tag{31}$$

$$\begin{aligned} &= \ln \sum_{m=0}^{M} \exp\{\max_{d_{it}} \max_{0 \le C_{it} \le I_{it}+Q_{it}} U\left(C_{it}\right) + \psi_{i,0}\left(Y_{it} - P_{it} - PC_{it} - IC_{it}\right) \\ &\quad + \delta \int V\left(x_{it+1}, \varepsilon_{it+1}\right) dG_{\varepsilon} dG_{P_{t+1}|P_{t,\tau}}\} \end{aligned} \tag{32}$$

Define $W\left(s_{it}\right)$ as the indirect utility function conditional on purchasing SKU m

$$\begin{aligned} W\left(s_{it}\right) &:= \max_{0 \le C_{it} \le I_{it}+Q_{it}} U\left(C_{it}\right) + \psi_{i,0}\left(Y_{it} - P_{it} - PC_{it} - IC_{it}\right) \\ &\quad + \delta \int \bar{V}\left(x_{it+1}\right) dG_{P_{t+1}|P_{t,\tau}} \end{aligned} \tag{33}$$

Then substituting eq (33) into eq (33)

$$\bar{V}\left(x_{it}\right)=\ln\sum_{m=0}^{M}\exp\left\{W\left(s_{it}\right)\right\} \tag{34}$$

Rolling forward eq (34) next period's integrated value function

$$\bar{V}\left(x_{it+1}\right) \;=\; \ln\sum_{m=0}^{M}\exp\left\{W\left(s_{it}\right)\right\} \tag{35}$$

Substituting eq (35) into eq (33) yields a Bellman Equation,

$$\begin{aligned} W\left(s_{it}\right) \;:=\; & \max_{0\leq C_{it}\leq I_{it}+Q_{it}} U\left(C_{it}\right)+\psi_{i,0}\left(Y_{it}-P_{it}-PC_{it}-IC_{it}\right) \\ & +\delta\int\ln\sum_{m=0}^{M}\exp\left\{W\left(s_{it+1}\right)\right\}dG_{P_{t+1}|P_{t,\tau}} \end{aligned} \tag{36}$$

Finally, I show that the household's discrete choice problem is to choose the largest indirect utility function once SKU specifics are realised

$$\begin{aligned} V\left(x_{it},\varepsilon_{it}\right) \;=\; & \max_{d_{it}}\max_{0\leq C_{it}\leq I_{it}+Q_{it}} U\left(C_{it}\right)+\psi_{i,0}\left(Y_{it}-P_{it}-PC_{it}-IC_{it}\right) \\ & +\sum_{m=0}^{M}d_{imt}\varepsilon_{imt}+\delta\int V\left(x_{it+1},\varepsilon_{it+1}\right)dG_{\varepsilon}dG_{P_{t+1}|P_{t,\tau}} \qquad (37) \\ =\; & \max_{d_{it}}\max_{0\leq C_{it}\leq I_{it}+Q_{it}} U\left(C_{it}\right)+\psi_{i,0}\left(Y_{it}-P_{it}-PC_{it}-IC_{it}\right) \\ & +\sum_{m=0}^{M}d_{imt}\varepsilon_{imt}+\delta\int\bar{V}\left(x_{it+1}\right)dG_{P_{t+1}|P_{t,\tau}} \qquad (38) \\ =\; & \max_{d_{it}}W\left(s_{it}\right)+\sum_{m=0}^{M}d_{imt}\varepsilon_{imt} \qquad (39) \end{aligned}$$

where the first line is eq (30). Eq (35) is substituted into the second line. Finally, I substitute in eq (36) and represent ε element-wise. This gives the desired expression.

Annex C: Estimation algorithm

At the start of iteration k of estimation algorithm, the value of structural parameters is θ_k and the current state of the solution to the dynamic program is given by the parameters r_k.

Denote the objective function at the beginning of the iteration from simulation of the model to calculate the moments at θ_k and r_k as

$$Q(\theta_k, r_k) := g(\theta_k, r_k)^\top \Sigma^{-1} g(\theta_k, r_k)$$

From the proposal density for the structural parameters I draw $\tilde{\theta}_k$ and simulate the model to compute a new objective function, $Q(\tilde{\theta}_k, r_k)$. Then, to decide whether accept or reject the draw of the structural parameters $Q(\tilde{\theta}_k, r_k)$ is compared to $Q(\theta_k, r_k)$. If the draw is accepted, set $\theta_{k+1} = \tilde{\theta}_k$. If not, I leave the structural parameters unchanged.

Next I do a single iteration of the λ-policy iteration algorithm at θ_{k+1}. Applying the policy evaluation and policy improvement steps described in Section 3.6 yields r_{k+1}. To prepare the algorithm for the next iteration, the objective function needs to be recalculated as $Q(\theta_{k+1}, r_{k+1})$. This is used in the accept-reject decision for the new structural parameter draws at r_{k+1}.

Conducting three separate simulations at each iteration is likely to be computationally burdensome. Clearly the simulation used to evaluate the draw is unavoidable and updating the ADP at each iteration is highly advisable. However, as highlighted by Imai et al. (2009), even if the objective function is not re-evaluated after the λ-PI step, the existing approximation to the objective function may have a good enough for the purpose of deciding whether to accept or reject the next structural parameters draw in the next iteration.

Following the suggested approach in Imai et al. (2009) a third simulation is conducted if the existing objective function is likely to be a poor approximation to $Q(\theta_{k+1}, r_{k+1})$. If $\tilde{\theta}_k$ is accepted then there is no need to re-simulate. The value of the objective function used in the structural parameter update can be used as an approximation; $Q(\theta_{k+1}, r_{k+1}) \approx Q(\theta_{k+1}, r_k)$.

If the draw is rejected the quality of the approximation to $Q(\theta_{k+1}, r_{k+1})$ may depend on how many iterations have passed since it was last updated. If a draw has been accepted in last $i \leq \bar{n}$ iterations, the objective function is left unchanged. That is, $Q(\theta_{k+1}, r_{k+1}) \approx Q(\theta_k, r_k)$ for $1 \leq i < \bar{n}$. After $i > \bar{n}$ successive rejections, after the λ-policy iteration step in iteration k, the objective function is re-evaluated at (θ_{k+1}, r_{k+1}). In the algorithm, if there is no update in the last five iterations, the objective function is re-evaluated.[52]

Next, is the adaptive element of the estimation algorithm. First, using the last 100 accepted parameter draws, the mean and covariance of the target density are calculated. The estimated moments of the target density pools parameter estimates across chains.

Then, the chain specific proposal covariance scaling factor is exponentially smoothed using the probability that most recent draw of the structural parameters is accepted. If the probability of acceptance, is greater (lower) than the target acceptance rate of 0.234, the scaling factor increases (decreases).

Next, using the ratio of the temperature scaled objective functions, chains swap probabilities are calculated. If the swap probability exceeds a draw from a uniform distribution, structural and value function parameters are swapped across chains.

[52]Given that the adaptive Markov Chain MC targets an accepted draw every 234 per 1000 draws, a 'natural' update occurs approximately once every 5 draws.

Finally, using the ratios of the temperature scaled objective functions in adjacent chains, the temperatures are updated ready for use in the next iteration of the estimation algorithm (see Łącki and Miasojedow (2015) for details). Łącki and Miasojedow (2015) also allow for a chain pruning phase. Because I only use a maximum of three chains, this phase is not included in the estimation algorithm used in this paper.

Annex D: Empirical Results

D.1 Interactive fixed effects model

Table 7: Price Model: interactive fixed effects model

		R=1	R=2	R=3
Formats:	Powder	-0.219	-0.449	-1.812
		(0.057)	(0.119)	(2.027)
	Liquid	-0.234	-0.709	-0.027
		(0.107)	(0.148)	(0.832)
	Tablets	-0.137	-0.574	-1.265
		(0.097)	(0.147)	(1.433)
	Gel	-0.396	-0.589	-1.822
		(0.062)	(0.119)	(1.905)
Brands:	Brand B	-0.143	0.068	0.673
		(0.073)	(0.110)	(1.190)
	Brand C	-0.463	-0.539	0.388
		(0.081)	(0.106)	(1.268)
	Brand D	0.120	0.277	0.795
		(0.057)	(0.106)	(0.916)
	Brand E	-0.312	-0.688	-1.693
		(0.093)	(0.170)	(1.855)
	Brand F	-0.712	-1.032	-1.436
		(0.089)	(0.175)	(1.229)
	PL	-1.020	-1.069	-1.466
		(0.090)	(0.112)	(0.964)
	Others	-0.828	0.377	-8.487
		(0.264)	(0.243)	(8.971)
Other Chars.:	Multipack	0.18	0.007	0.024
		(0.049)	(0.118)	(0.156)
	Washes	-2.922	-0.383	-0.465
		(0.975)	(1.708)	(2.316)
	Washes^2	2.096	-1.106	-1.050
		(1.758)	(2.439)	(3.261)
	Washes^3	-0.684	0.808	0.815
		(1.016)	(1.267)	(1.666)
Constant		0.789	0.689	1.551
		(0.138)	(0.196)	(1.363)
R^2: Overall		0.674	0.710	0.728
R^2: Within		0.364	0.132	0.061
Num. Obs.		14,927	14,927	14,927

Annex E: Inventory Proxy

Construct proxy for inventory for household h whose first period in the data is T_0 and last is T.

1. Calculate average consumption, $\bar{C}$, over the sample

$$\bar{C} = \frac{\sum_{t=T_0}^{T} Q_t}{T - T_0 + 1}$$

   where Q_t is the number of washes purchased in period t

2. Set inventory of 0 immediately prior to first purchase,

$$I_{T_{-1}} = 0$$

3. Then for $t = T_0, \ldots, T$ calculate inventory before purchase are made

$$I_t = \max\left\{0, I_{t-1} + Q_{t-1} - \bar{C}\right\}$$

4. The first 12 periods are omitted to reduce dependence on initial inventory.

Annex F: Additional Results

Figure 14: Re-purchase probabilities: formats