Learning to Set Prices in the Washington State Liquor Market

Yufeng Huang, Paul B. Ellickson and Mitchell J. Lovett*

June 26, 2018

[Preliminary and comments welcome!]

Abstract

How quickly do new entrants learn about demand and adapt to the new market? We study retailer pricing in the Washington State liquor market where its privatization leads to existing grocery chains entering this market for the first time. We document large price changes across a broad range of products and provide novel evidence showing that these changes result from retailers learning about demand: prices absorb realized demand shocks and adjust to better reflect demand primitives. We then estimate a structural model with minimum assumptions on the optimality of observed prices. Comparing against the full-information optimal prices implied by the model, we find that (1) learning exists years after entering the new market, (2) upon entry, limited demand information causes 11% lower profit compared to full-information, and (3) there are sizable heterogeneity between retailers in initial information sets and in learning rates.

*University of Rochester Simon Business School. We thank comments and suggestions from Kristina Brecko, Chris Conlon, Ulrich Doraszelski, Ronald Goettler, Brett Gordon, Avery Haviv, Przemyslaw Jeziorski, Sarah Moshary, Simha Mummilaneni, Xiliang Lin, Marc Rysman, Thomas Wollman, Chenyu Yang, and Hongsong Zhang, as well as seminar and conference participants in Northwestern University, SHUFE IO Conference, University of Chicago, and University of Rochester. Our empirical results are derived based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are those of the researchers and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.
1 Introduction

“Pricing is a big question mark, for everyone entering the spirits business in Washing-
ton... I sure don’t know what we’ll charge the consumer. There is going to be a lot of scrambling...”

– Alan Johnson, CEO of BevMo!

Firms face, perhaps, no more tenuous a knowledge position than when entering a new market. Existing knowledge needs to be adapted to new realities, and the firm will likely need to update its beliefs as information about the new reality is revealed. Because the different knowledge positions of firms affect their decisions, how quickly firms adapt to the new environment can potentially have a profound influence on firm performance (Noble and Gruca, 1999; Bloom et al., 2017). Furthermore, because firm behavior plays a vital role in determining equilibrium market outcomes, empirically studying firms’ adaptation to new market environments is vital to policy decisions – such as merger (Björnerstedt and Verboven, 2016), taxation (Seiler et al., 2018), market privatization (Miravete et al., 2017) – which often involve drastic market shifts.

Recent empirical evidence documents that, after entering a market, firms initially behave in ways that appear suboptimal. Doraszelski et al. (2016) find that firms who recently entered a new market change their prices more frequently and argue that they do not seem to follow a well-defined equilibrium strategy. While firms may initially lack the knowledge they need to make optimal choices, we should expect decisions and performance would to over time. The existing evidence on such firm learning has taken an internal focus on either organizational aspects (e.g., Hurley and Hult, 1998) or about productivity or production experience curves (Argote and Epple, 1990; Benkard, 2000; Covert 2015). However, the literature is relatively silent about learning about demand – the focus of the current research – with a few notable exceptions (Hitsch 2006; Jeon, 2017). We build on this literature and empirically study the evolution of firm strategies when they enter a new market, and in which ways firms differ in learning.

To study learning about demand in the field is difficult because most settings where such learning is likely to play a substantial role present many confounds. Learning after disruptive innovations involves a complex innovation game between competitors (and potentially partners); learning when entering a new market involves reacting to other entrants, evolving customer tastes, new firm routines, and new market information systems. Typically, the presence of many moving parts makes it difficult, if not impossible, to isolate learning about demand from a myriad of other possibly more important factors. We identify a context for our study that provides an almost ideal, lab-like setting to examine learning about demand.

We study pricing decisions made by established grocery retailers upon entering the liquor market in Washington State after privatization of the market. In Washington, existing products were sold by a state-owned chain, who committed to following a fixed-percent markup pricing rule and did not set prices based on demand conditions. The well-anticipated privatization took effect in June 2012 and allowed larger retailers to obtain licenses to sell liquor. This setting has several features that simplify our inference problem: customers with stable preferences for liquor, retailers with established positions and stable customer bases for which liquor does not appear to drive store visits, and a managerial decision, pricing, that has established routines that can directly extend to the new category. These aspects of the setting allow our inquiry to focus on the most critical uncertainty that retailers faced in setting prices – consumer demand for products. Our study focuses on whether, after entering the liquor market, these chains learn about demand over time and improve their pricing decisions.

We document large and heterogeneous price movements in the first 2 years after the privatization of liquor sales. While average prices moved relatively little beyond the first few months, most of the products changed their prices by at least 5% (up or down) and 15% of products changed their prices by more than 15%. We present two pieces of descriptive evidence that suggest these price movements are due to retailers’ learning about demand. First, we show that retail prices for a product respond to lagged demand shocks for the same product and that the rate of this response declines over time. This pattern is consistent with retailers learning about consumer preferences.
for specific products—i.e., they respond to sales shocks as new information, but the influence of this new information decreases as the retailer gains more information about preferences. Second, novel to the literature, we show that, across products, the correlation between observed prices and average sales increases over time. This suggests that, as retailers learn, they identify the products that consumers are willing to pay more (less) for and set higher (lower) prices correspondingly. We also provide evidence that these price movements are not reflective of a number of other potential explanations including changes in retail outlets, assortment, competitive environment, consumer tastes or price search.

We then ask more formally whether firms learn to set prices better as a result of these behaviors. To do so, we need a normative model to compare the observed prices against. To form this comparison, we estimate a structural model of demand and cost primitives, imposing minimal assumptions on the optimality of the retailers’ pricing decisions. We estimate a random coefficient demand model incorporating standard aggregate- and micro-level moments. Our model estimates and outputs accord well with existing evidence from related liquor markets (Miravete et al., 2017). We then use the model to simulate the normative benchmark of optimal, full-information pricing.

We present three conclusions. First, retailers do learn about demand over time; prices become more similar to the optimal full-information levels. Second, early in the market, prices are set to sub-optimal levels and the lack of optimality corresponds to as much as a 11% loss in gross profit compared to the (unattainable) perfect information case. Third, we find significant across-chain heterogeneity in their initial knowledge positions and the pace at which they learn. In particular, retailers that operate only in Washington have worse initial price positions, but learn more rapidly, whereas retailers that also operate outside of Washington start with profits that are closer to optimal, but take longer to improve pricing decisions to further enhance profits. We provide further descriptive evidence that the latter retailers seem to be constrained, in line with the literature on national retail prices (Adams and Williams, 2017; DellaVigna and Gentzkow, 2017; Hitsch et al., 2017) but differ from the studies on heterogeneity in firm strategies, which find that large firms behave closer to model predictions (Hortacsu et al., 2017).
Our main contributions are two-fold. First, we provide new evidence on firm learning and the impact of related practices on profits. We show that retailer strategies respond to past demand shocks in a way consistent with learning about demand for specific products. We further establish that retailers set prices that increasingly capture demand, suggesting that they obtain increasingly precise information about its products. To our knowledge, the second piece of evidence is novel to the literature. The closest related paper is Doraszelski et al. (2016), who show that after the market for frequency response opens (in the UK electricity system), prices appear random at first and evolve over time to patterns that are close to a stationary equilibrium. They focus on learning about equilibrium play whereas our paper focuses on retailers learning about underlying demand conditions. Our paper is also related to Hitsch (2006), Jeon (2017) and Covert (2015). Hitsch (2006) documents that, in the ready-to-eat cereal market, many products remain in the market for a long time despite making low sales, and estimates a structural model in which forward-looking manufacturers learn about the demand for their products. Related to Hitsch (2006), Jeon (2017) studies dynamic investment decisions of container shippers as they learn about stochastic demand in the container shipping industry. Given data on shipments and investment costs, she estimates primitives of a learning model and finds that producer welfare would increase with full information. Covert (2015) investigates firms’ learning about shale productivity in the Hydraulic fracturing industry. He shows that firms respond to public reports about production output and adjust their production strategies, and finds that firms overweight own signals and incorporate public signals sub-optimally.

Second, we contribute to the discussion about heterogeneous management practices within firms. There are broadly two approaches in this literature. In one approach, Noble and Gruca (1999) and Bloom et al. (2017) base their evidence on surveys and show that there is heterogeneity in firms’ reported business practices. Bloom et al. (2017), among others, show that such differences explain a sizable fraction of the productivity differences across firms. In the other approach, Goldfarb and Xiao (2011) and Hortacsu et al. (2017) show that there is heterogeneity in observed strategies. Whereas the strategies for some firms better fit an economic model (e.g. in Hortacsu et
al., larger firms behave closer to model predictions), the behavior of other firms is more difficult
to rationalize. Our approach lies closer to the latter, and we add to this literature by showing that
pricing strategies can be sub-optimal when a firm is inexperienced – and more importantly, this
lack of experience is not only an outcome of lack of information but rather lack of understanding
of existing information.

Secondarily, we also add evidence to recent descriptive studies of retail pricing. Chintagunta
et al. (2003), Adams and Williams (2017), DellaVigna and Gentzkow (2017) and Hitsch et al.
(2017) document that retail prices are geographically uniform and discuss the extent to which they
are set sub-optimally by retailers who ignore heterogeneity across markets. We confirm the same
case in Washington State liquor market. In addition, we speak to this literature in our evidence on
heterogeneity in learning: We show that local Washington retailers converge quickly to state-level
optimal prices, whereas multi-state retailers display sub-optimal pricing behavior that is reminis-
cent of cross-state promotion strategies.

Finally, our paper also contributes to the research and evidence in recent studies about the
retail liquor industry. Seo (2016) studies consumer store choice before and after the privatiza-
tion of Washington State liquor market and quantifies the welfare impact of one-stop shopping.
Illanes and Moshary (2017) leverage Washington State’s 10,000 square-foot minimum required
retail space for liquor vendors, as a regression-discontinuity design, and study the effect of en-
try on prices and product assortments. Conlon and Rao (2015), Aguirregabiria et al. (2016), and
study the “post-and-hold” policy as a collusive instrument for the wholesalers and investigate the
welfare improvements (and redistribution) of an alternative tax policy. Aguirregabiria et al. (2016)
investigate counterfactual regulation, tax, or competition regimes in the Ontario wine market, high-
lighting the importance of spatial differentiation. Miravete et al. (2017) study the welfare impact
of state-imposed constant retail markup and find that the single-markup policy decreases but also
re-distributes consumer welfare.

Section 2 briefly describes the privatization of the Washington liquor market. Section 3 presents
the data, and Section 4 provides key descriptive evidence about retailer learning. Section 5 then estimates a structural model and backs out demand and costs. After the structural estimation, Section 6 describes firms’ learning process from the structural estimates. Section 7 concludes the paper.

2 Privatization of liquor sales in Washington State

Prior to June 2012, off-premise liquor sales in Washington State was monopolized by a state-owned chain, “Liquor and Wine”, that was operated by the Washington State Liquor Control Board (WSLCB). The state-owned chain committed to set prices at a 51.9% markup over the wholesale price. In Fiscal 2011, WSLCB directly operated 166 stores in cities and contracted with private owners who operated 162 stores in rural areas. Total sales revenue for alcoholic beverages amounted to $888 million in 2011.\(^2\)

On November 8, 2011, Initiative 1183, which mandated that the Washington State liquor business be privatized, was passed with 59% of voter in favor. I-1183 mandated 1) that WSLCB must exit the liquor business before June 1, and auction off its inventories; 2) that private retailers are eligible to carry and sell liquor products upon obtaining a license, starting from June 1, 2012; 3) that private retailers must pay 17% of liquor revenue to the state as part of the licensing fee; and 4) that retail off-premise sales taxes will increase from 10% to 20.5% (excise tax, at $3.77 per liter, stays unchanged). In particular, the Initiative requires that retail licenses can only be issued if the store has at least 10,000 square feet of floor size. Before I-1183, two other initiatives to privatize the retail liquor market were voted down in 2010.

After the privatization of liquor sales, retail licenses were issued from March 1, 2012, and we observe that grocery stores and other retail chains (who do not specialize in liquor sales) immediately became dominant players starting from June 1. That grocery retailers dominated the market

---

\(^2\)This percentage markup was 39.2% and was increased to 51.9% on August 1, 2009. Source: WSLCB press release, May 6, 2009.

is consistent with Article (3)(b) in I-1183, which prioritizes the issuance of retail liquor licenses for “existing grocery premises licensed to sell beer and/or wine”. Also, the press reported that many of the state stores remained in business but are operated by independent owners.

A report by the State shows that post-privatization, total sales volume increased by about 20%, average prices increased by about 8%, and the total number of liquor-selling stores increased by 3.27 times (Washington-State, 2015). Similar numbers are reported by Seo (2016) and Illanes and Moshary (2017).

3 Data and summary statistics

3.1 Data and sub-sample

Our main data source is the Nielsen Retail Measurement Scan (RMS) Dataset.\textsuperscript{4} For participating chains, the Nielsen RMS dataset contains information about price, quantity sold, and feature and display, for a large set of consumer packed goods, recorded at the store-UPC-week level. For our study, we focus on the liquor category in Washington State, from June 2012 (the month of privatization) to December 2016. To limit the complexity of our analysis (especially in the structural model), we further focus on the broad whisky category, which consists of whisky, bourbon, scotch and rye, a set of fairly closely substitutable products. This sampling procedure yields a dataset of 6,417,108 observations at the UPC-retailer-store-week level.\textsuperscript{5} In particular, the sample contains 719 unique UPCs (UPCs as product name - size combinations; there are 617 unique product names), 660 stores from 7 retail chains,\textsuperscript{6} and 240 weeks between June 2, 2012 and December 24, 2016.

We further restrict our attention to a smaller subsample through the following two steps. First,

\textsuperscript{4}We thank the Kilts Center for access to the data. Nielsen retains copyright of the data: Copyright © 2018 The Nielsen Company (US), LLC. All Rights Reserved.

\textsuperscript{5}Only for Washington State.

\textsuperscript{6}The sample we discuss has already excluded two grocery retailers that entered in small scale in some parts the market in 2016 (24,225 observations, 0.3% of the sample size) and two mass merchandisers that only have 238 observations (0.0006% of the sample size). These observations are anomalous and so we dropped them.
we focus on stores that sell a positive quantity in at least 95% of all weeks. This step selects 594 out of 660 stores, and removes 6.2% observations from the overall sample. We note that most of the cases in which quantities go to zero are not due to the decision to carry or drop liquor, but rather are due to entry or exit of the store as a whole. Therefore, we condition on the set of stable stores.\(^7\)

Second, we focus on “core assortments” defined as those, within a given retailer, that first appear before December 2012 and last appear after March 2016. This means that we condition on a set of products that the retailer carries approximately from the beginning to the end of the sample, abstracting away from new products or discontinued products. This step selects 276 out of 719 UPCs in the full sample. Although it seems that most of the UPCs are eliminated in this step, the products we drop out in this step only account for 16.1% of the sample size of stable stores and 11.7% of the total revenue from these stable stores. We discuss the choice of focusing on these core products in more detail in Section 3.3.2. After these sample selection steps, our sample contains 5,053,838 observations, from 276 products, 7 retailers, and 594 stores.

Washington is the first state to privatize its liquor sales. In some of our descriptive analyses, we compare Washington to other states that have allowed retail grocery liquor sales for as much as several decades. We identify fifteen such states in the Nielsen RMS data and use these as a kind of “placebo test.”\(^8\) In this auxiliary sample, retailers have generally been in the liquor business for a long time in comparatively stable market conditions, so that one does not expect them to exhibit strong learning about demand. We therefore contrast our evidence about retailer learning in Washington against evidence in those states.

Finally, we also use the Nielsen Consumer Panel Data between December 2009 and December 2016. We use this dataset to examine consumer behavior in greater detail and to construct micro-moments (Petrin, 2002) for use in structural demand estimation. Within the panel data, there are 2,952 households residing in Washington State, 1,177 of them ever purchased liquor in the

---

\(^7\)By the account of Section 2, our sample contains 55% of all the liquor-selling stores in the state.

\(^8\)These states are, in descending order of total liquor sales volume within our data: California, Arizona, Louisiana, Texas, New Mexico, Nevada, Nebraska, Wyoming, South Dakota, Colorado, Arkansas, Delaware, Maryland, North Dakota, and Washington, D.C.
sample period, and 498 households purchased at least five times (in all stores). Hence, because the household panel data contain very infrequent liquor purchases, we rely on the RMS data for our main analyses.

3.2 Price changes after the privatization of liquor sales

3.2.1 At which geographic level are prices set?

As our focus is on learning about demand, it is important to establish where that learning might manifest itself and the degree of price variation observed over time. Consistent with Hitsch et al. (2017) and DellaVigna and Gentzkow (2017) we find that for a given product at a given point in time, there is little price variation within a chain across stores and markets (see Appendix A for details). However, across retailers and over time, there are sizable price differences. Therefore, we treat pricing decisions as if they are made on the retail chain level and investigate changes in these prices over time (and the extent to which such changes are driven by learning about demand).

3.2.2 How much do prices change over time?

We first note that, across product-retailers, the median price drop from $23.7 to $21.8 in the first six months after privatization. After this price decline, the median price stays relatively stable.

Second, we seek to quantify the magnitude of price changes by characterizing the distribution of changes relative to each product’s initial price. For each product-retailer, we first calculate the percentage changes in its prices (averaged across markets) relative to the initial price, \( \frac{\bar{p}_{jrt} - \bar{p}_{jrt_0}}{\bar{p}_{jrt_0}} \). We then plot the median, 25th/75th percentile, and 10th/90th percentiles of the normalized prices over time. Figure 1 shows the changes of these quantiles: In Washington State, prices underwent dramatic changes and the implied price paths are heterogeneous. For example, in mid-2014, 25% of products are priced at least 8.6% below their starting price (at the privatization), and 25% of products are priced at least 6.3% above their starting price. In addition, much of the price movements happen in the first 2 years, with the distribution of price changes...
becoming stable afterwards. Although this is not direct evidence of learning, such dramatic price changes are consistent with retailers learning about demand and adjust (the distribution of) prices.

Further, we do note that in the final quarter of the sample, there is a large decrease of over 5 percentage points. We do not have an explanation of the phenomenon within our model, and thus for our main analyses, we end the sample period prior to this price drop.

Appendix Figure 16 presents another way to visualize the price paths, plotting the relative prices (defined above) over time for all products separately.

### 3.2.3 Changes in the overall price level and promotion policy

It is also important to consider whether promotion policies evolved over the sample period, perhaps contributing to or explaining the price changes. To this end, we first define regular price as the 90th percentile of prices for a given store-product in a given quarter, and define promotion instance as when the distance between list price and regular price is beyond 5% below the regular price (and promotion depth accordingly). In Washington, promotion as defined occurs at a frequency of 26%,

![Figure 1](image-url)
and average promotion depth is 9% of regular price when the product is on promotion.

Focusing on the time trend in list price, regular price, promotion frequency and promotion depth, we find that list price and regular price follow similar trends, and there is no meaningful variation in promotion frequency or depth over time. Promotion frequency varies over time in the scale of -5% to +10%, and such variations appear to be cyclical. Promotion depth increases (i.e. promoted price decrease) by 1% at the end of the sample, compared to mid-2012. Given the average promotion frequency, such promotion depth change only explains less than 1% of price change. Thus, it appears that promotional policy does not have systematic changes that need to be characterized separately. We present the results in Appendix Figure 18.

3.3 Do price changes reflect phenomena other than learning about demand?

In this section, we examine a few factors (other than prices) that might change because of the privatization of liquor sales. Our aim is to rule these out as alternative dimensions on which firms might be updating their beliefs.

3.3.1 Changes in the set of stores that carry liquor products

One might expect retailers to make decisions about which stores to carry liquor and for such decisions to change as they learn about demand. We examine whether the set of stores selling liquor changes over time and rule out this conjecture. We compare the number of distinct stores selling grocery items for a given chain, against the number of stores selling liquor for the same chain. Appendix Figure 14 report these numbers over time and by chain. For the set of chains in the liquor market, almost all stores carry liquor as long as they remain in business. We conclude that which store to carry liquor was not a relevant decision for the chain/store managers.\(^9\)

\(^9\)One might be concerned that stores that sell little quantity might be mis-recorded as not carrying liquor, as liquor is a slow-moving category. We collect data on the identity of off-premise liquor license holders by year from the Washington State Liquor Control Board, cross-check the number of distinct license holders by retailer by year, and find that the number of license exactly match the number of stores in the Nielsen data. This additional check indicates that there is no measurement error in the store identity.
3.3.2 Changes in liquor product assortments

Alternatively, one might expect that retailers make decisions about which assortments to carry and that such decisions could also be an outcome of learning. We examine whether the decision of which products to carry changes over time and gauge the overall magnitude of assortment changes. As mentioned in Section 3.1, among the 730 products (as product name - size combinations, sizes can be 375ml, 750ml or 1750ml), only 276 products are “core” in the sense that they are carried by the retailer during 2012 to 2016. The “non-core” products represent 16.1% of the sample and account for 11.7% of the overall revenue. Appendix Figure 15 shows these core products account for most of the revenue over time.

Given the small revenue shares for the set of products that vary over time, we focus our inquiry around the core products. That is, we conclude that, although firms may adjust assortment as they learn about demand, this decision appears relevant only for low-volume products.\footnote{Choosing to discontinue unexpectedly low-volume products is discussed in Hitsch (2006).}

3.3.3 Changes to the market structure

One might expect that the market structure might change over time as new entrants come to the market and compete against the initial set of entrants. Figure 2 shows that the total revenue and revenue composition among the seven focal retailers are stable over time: it seems that until the first half of 2016, the market structure is stable as each retailer occupies a stable share of the market.

Nielsen RMS data does not cover all retailers in the market. The Homescan data cover, for a small set of consumers, their expenditure in all retailers. We use these data to measure the revenue of the six retailers in our main data relative to the rest of the market. We find that after the market is privatized, the six retailers collectively take 37% of the total revenue on average. In addition, their revenue share in the market is stable over time, as shown in Appendix Figure 17. These descriptive evidence suggest that market structure does not change within our sample period.
Figure 2: Whisky sales revenue decomposition, across retailers

Notes: Decomposition of liquor sales revenue across six focal retailers in Washington State, focusing on the set of core products.

3.3.4 Behavior of liquor shoppers

Finally, one might expect that consumer behavior changes over time after the privatization of the market. In this section, we present some descriptive analysis focusing on the behavior of liquor shoppers, and examine the scope for such consumer learning. We group this evidence in Figure 3.

First, after privatization, retailers may view liquor prices as a means to attract new customers. However, as we previously discussed, prices were actually higher after privatization than before and generally fall over time. Second, consumers may discover the liquor category over time as a result of in-store marketing that retailers are able to target to existing shoppers who otherwise would not have encountered such marketing. In fact, privatization of the market was heavily publicized for two years prior to the launch. In the top-left panel, we show that household expenditure on liquor increased immediately after the event and stayed constant. Hence, beyond the immediate increase in liquor sales, there is no evidence of attracting new customers or expanding customer purchases over time.

Second, even if the availability of the liquor category is well known in advance, consumers
might have meaningful shifts in preferences after deregulation and over time. For example, such changes and consumer learning might arise from assortment changes or in-store marketing that brings new varieties to the consumers attention. We examine whether different measures of store and product variety change over time after the privatization. The bottom four plots of Figure 3 show that, across four different measures– (1) the number of distinct chains a consumer visits, (2) the number of distinct sub-categories or product types (e.g. scotch or bourbon) the consumer purchases, (3) the number of different brands, and (3) the quantity of liquor products purchased per month, conditional on purchase–no identifiable time patterns exist after the policy change. Again, beyond the noticeable increase in purchases immediately at privatization, these descriptive patterns do not support the conjecture that consumer preferences change following privatization.

Lastly, because privatization allows competition among retailers, consumers could take advantage of this expanded retailer set to price shop for liquor products across stores. We address this concern in three ways. First, we show in the top-right panel of Figure 3 that liquor expenditure is only a small fraction of total grocery expenditures even in the trips where consumers buy liquor. This finding suggests that when consumers buy liquor at a grocery store, it is rarely the main purpose of the trip. Second, we also verify that there is no discernible systematic changes in the store choice patterns of consumers, before and after liquor privatization. Given that some but not all grocery retailers carry liquor post-privatization, such pattern suggests that consumers do not start shopping at new grocery stores in order to price shop for liquor. Third, in Appendix B, we show that sales of a product does not decrease if other retailers in the same market carry the product or promote the product. These evidence further suggest that substitution between chains for a liquor product is not detectable in our sample. Therefore, is it a reasonable approximation that grocery retailers set liquor prices for existing customers as if they are a monopolist on this customer base.

11We also note that expenditure and other household choices, measured by the Homescan data, fall slightly in the last 1-2 years. In contrast, the point-of-sales data (RMS) do not show such a pattern. We examine the changes in household demographics and find a slight decrease of income among active shoppers. Thus, we do not interpret the slight decrease in household behavior at the end of the sample as representative of the market.
Figure 3: Descriptive evidence for consumer behavior in the WA liquor market

Notes: These panels present additional descriptive evidence from the household panel. In the top-left, we present changes in liquor expenditure before and after the privatization. For this panel, we estimate linear regression of log liquor expenditure, \( \log(\text{expd} + 1) \), on a set of half-year dummies, for all consumers in Washington state in the household panel. The figure reports these regression coefficients. We re-defined half-years to “December to May” and “June to November” in order to align with the timing of the policy change. In the top-right, we present share of liquor expenditure in a trip to grocery store. The bottom panels are additional measures of varieties: the number of distinct chains, liquor product types, brands, and bottles of liquor per month, for trips with liquor purchases.
4 Learning about demand: empirical evidence

In this section, we provide two key pieces of evidence, showing that retailers in Washington State learn about demand for liquor and adjust their prices. First, we show that prices adjust according to past demand shocks when retailers start selling liquor in Washington, and that, over time, they adjust less and less to such shocks. This is consistent with a rational learning model where a retailer who does not know about demand will adjust prices according to temporary demand shocks, whereas the retailer who later knows demand will not adjust prices to these shocks. Second, we show that the price of a product increasingly reflects the demand level of this product (i.e., intercept in the demand function). This implies that retailers gain knowledge about which product sells better in Washington and are able to adjust their prices to reflect such knowledge.

Although our main focus is to test for learning in Washington, we present the same set of evidence in other states as a “placebo” test. In other states, retailers have been selling liquor products way before June 2012. Therefore, we expect that our evidence for learning should not appear in these placebo tests.

4.1 Do prices adjust based on past demand shocks?

If a retailer learns about demand by observing (noisy) realizations of demand shocks, she will initially adjust prices according to those shocks, but later (as she has already learned) cease doing so. We test whether retailers adjust prices according to demand shocks when they enter the market and whether they do so to a lesser extent as they gain experience.

We first aggregate our data to the product-retailer-month level to reduce the effect of temporary price promotions and to work at the geographical aggregation where prices are set. Denote \( j \) as a product, \( r \) as a retailer and \( t \) as a month, we estimate a linear model of current price on 1-month
lagged quantity, controlling for current quantity and lagged prices:

\[
\log(p_{jrt}) = \beta_t \log(q_{jrt-1}) + \\
\rho \log(p_{jrt-1}) + \alpha^{-1} \log(q_{jrt}) + \psi_{jr} + \phi_t + \eta_{jrt}.
\]  

(1)

Our key parameter of interest is the sensitivity of the current price to \(q_{jrt-1}\), which is the units sold for product \(j\) by retailer \(r\) in the previous month, \(t - 1\). We allow \(\beta_t\) to take a different value for each half-year. Learning would predict that these sensitivities should be initially positive, but then decline in magnitude over time. We include product-retailer fixed effects, \(\psi_{jr}\), that capture stable product-retailer characteristics that could affect price levels, time fixed effects, \(\phi_t\), that capture common variations price across time, past logged prices (with parameter, \(\rho\)) which captures serial correlation in price beyond these fixed effects, and logged current quantity (with parameter, \(\alpha^{-1}\)) that captures the correlation between current quantity and price,\(^{12}\) and can be approximately thought of as the inverse price elasticity.

For estimation, we take first difference to net out the fixed effects. After first differences, \(\Delta \log(p_{jrt-1})\) and \(\Delta \eta_{jrt}\) are mechanically correlated because of the first difference, and we correct for such correlation using Arellano and Bond (1991) instruments. To instrument the first lag of price, we use the third lagged price difference, \(\Delta \log(p_{jrt-3})\), which guards against potential additional serial correlation in \(\eta_{jrt}\) that might infect the second lag. Such serial correlation could potentially arise because retailers are unwilling to change price right after a price change due to reasons such as menu costs.

Table 1 presents these estimation results and Figure 4 plots the coefficients. We estimate and plot separately Washington and the placebo states where retailers are experienced in selling liquor. We check the first stage and find that the Arellano-Bond instruments are strong and the coefficients on control variables have plausible sign and magnitude. As main results, we find that in Washington, prices respond to the previous month sales quantity positively, but the effect decreases over

\(^{12}\)Current quantity is correlated with past quantity and thus cannot be omitted from the regression.
Figure 4: The response of current price to lagged quantity, Washington (left) and other states (right)

**Notes:** Estimates $\beta_\tau$ from Table 1, with 2-standard-error confidence intervals.

time. Just after the privatization of liquor sales, an increase from the mean of 10% in past sales increases prices in the next month by 0.16%. Four years later the responsiveness decreases by half to only 0.08%. Hence, retail prices incorporate past sales shocks but decreasingly so, consistent with retailers learning about demand by incorporating information contained in the realized sales quantities. This pattern for Washington State stands in sharp contrast to what we find for the states where retailers already had extensive experience in the liquor market. In these states, the responsiveness to sales shocks is much lower and all but one coefficient are insignificant. Hence, with greater experiences, prices do not respond to past sales shocks.\(^{13}\)

### 4.2 Do prices increasingly reflect demand?

One of the early motivations for price endogeneity bias is that prices are set in part to reflect the underlying quality of the product (Trajtenberg, 1989; Berry, 1994). As a result, if cross-sectional variation were used to estimate price sensitivity, the price sensitivity would be biased toward zero.

\(^{13}\)Except for the second half of 2012.
<table>
<thead>
<tr>
<th></th>
<th>WA: D.logprice</th>
<th>first stage</th>
<th>Ctrl: D.logprice</th>
<th>first stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>past month D.logprice</td>
<td>0.198***</td>
<td>-0.109***</td>
<td>0.182***</td>
<td>-0.067***</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.055)</td>
<td>(0.037)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>L3D.logprice</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>D.log quantity</td>
<td>-0.038***</td>
<td>-0.003***</td>
<td>-0.035***</td>
<td>-0.001***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>D.past month log quantity (2012.5)</td>
<td>0.016***</td>
<td>-0.052***</td>
<td>0.005***</td>
<td>-0.035***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2013)</td>
<td>0.013***</td>
<td>-0.048***</td>
<td>0.003*</td>
<td>-0.036***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2013.5)</td>
<td>0.014***</td>
<td>-0.045***</td>
<td>0.003</td>
<td>-0.034***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2014)</td>
<td>0.007**</td>
<td>-0.042***</td>
<td>0.001</td>
<td>-0.034***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2014.5)</td>
<td>0.008***</td>
<td>-0.036***</td>
<td>0.002</td>
<td>-0.031***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2015)</td>
<td>-0.001</td>
<td>-0.032***</td>
<td>0.002</td>
<td>-0.031***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2015.5)</td>
<td>0.005**</td>
<td>-0.028***</td>
<td>0.002</td>
<td>-0.030***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2016)</td>
<td>0.004*</td>
<td>-0.025***</td>
<td>0.001</td>
<td>-0.027***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>D.past month log quantity (2016.5)</td>
<td>0.008***</td>
<td>-0.018***</td>
<td>0.001</td>
<td>-0.017***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>constant</td>
<td>-0.039***</td>
<td>-0.016**</td>
<td>0.020***</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>yearmonth FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rsq.</td>
<td>0.002</td>
<td>0.122</td>
<td></td>
<td>0.054</td>
</tr>
<tr>
<td>obs.</td>
<td>3.3e+04</td>
<td>3.3e+04</td>
<td>2.2e+05</td>
<td>2.2e+05</td>
</tr>
</tbody>
</table>

**Notes:** Column 1 reports estimates of Equation (1), estimated using Washington data (first two columns) and compared against data from other states (last two columns). Lagged price difference is instrumented by the third lagged price difference, and Column 2 reports first stage results (dependent variable is lagged log price difference). Columns 3 and 4 estimates the same specification on other states. All standard errors are robust and clustered on the product-retailer level. *, **, and *** indicate significance at the 90, 95, and 99 percent. F-test for excluded variables in first stage: 392 for Washington, 912 for other states.
We extend this logic to the setting where the price-setter is learning about demand. We argue that if retailers are uncertain about demand initially but learn about it over time, prices should evolve to increasingly reflect demand as the learning process unfolds. Prices for products that have surprisingly high-demand (relative to the prior belief) will rise to reflect that high demand, and prices for surprisingly low-demand products will fall. The end result is that prices and the demand levels of products should become more highly correlated as the retailer learns about demand. Hence, learning should be reflected in a pattern of increasing correlations between prices and the underlying demand for the product.

This intuitive argument can be illustrated with a simple model has linear demand for product \( j \) in retailer \( r \) and month \( t \) determined by price and an intercept \( \delta_{jr} \),

\[
q_{jrt} = \delta_{jr} + \alpha p_{jrt}.
\] (2)

Assuming profit maximization and zero marginal costs, the optimal price would be set to \( p_{jrt} = -\frac{\delta_{jr}}{2\alpha} \). Imagine the slope \( \alpha \) is known to the retailer from serving these customers for many years in other categories, but the intercept \( \delta_{jr} \), which is specific to the new market, is not fully known.\(^{14}\)

As retailers learn about demand, their belief \( \delta_{jr} \) is closer to the truth, and so are the prices to the optimal, full-information ones. Hence, as retailers learn over time, the correlation between the prices and this underlying demand primitive would on average increase.

This observation and illustration motivates a new descriptive test of firm learning about demand. The concept is to estimate the cross-sectional relationship between prices and the sales for a product in each period and evaluate how this correlation changes over time. If learning occurs, we would expect the correlation to increase over time and otherwise be (potentially noisy, but) stable. To implement this test, we focus on quantity and price data on the product-retailer-week level. Pooling these data to half-year time windows (denoted \( \tau \)), we estimate the linear regression

\(^{14}\)Although this illustration assumes knowledge of price sensitivities and zero marginal costs, these are not necessary for the main conclusion to hold.
between quantity and price by ordinary least squares (OLS)

\[ q_{jrt} = \delta_{jrt} + \alpha_{r}p_{jrt} + \eta_{jrt}. \]  

(3)

With the retailer-week fixed effects $\delta_{rt}$ in Equation (3), the remaining variations left in the error term is only across-product, i.e. $\eta_{jrt} = \delta_{jr} + \xi_{jrt} - \bar{\delta}_{rt}$ (where $\xi_{jrt}$ represents idiosyncratic demand shocks). Thus, besides capturing the slope of demand, the regression coefficient $\alpha_{r}$ might also reflect retailer’s belief about product quality, through the covariance between $\delta_{jr}$ (left in the error term) and some belief about it (reflected by the observed prices). We report the estimated $\alpha_{r}$’s and their standard error for the Washington sample in Figure 5. We also examine retailers in other states as a placebo test: in that case, we control for retailer-state-week fixed effects in a similar regression as (3).

In Washington, we find that retail prices are increasingly correlated with demand intercepts, reflected by the way $\hat{\alpha}_{t}$ increases between 2012 and 2015. The relationship then stabilizes for the last two years. If the underlying price sensitivities are stable over time this finding suggests that retailers set prices with more and more information about the demand, so that the price sensitivities are biased more and more toward zero. In contrast to this pattern, we find that the other states exhibit no meaningful increases in $\hat{\alpha}_{t}$ and that it is instead stable over time. This finding is consistent with these experienced retailers exhibiting no meaningful learning about liquor demand. Both of the plots support the conjecture that retailers in Washington learned about demand after deregulation.

We note that we return to this concept for testing learning in section 6.1. In that section, we estimate our structural demand model parameters, which includes parameters for $\delta_{jr}$. We then correlate those parameter estimates with the observed prices to see whether this correlation increases over time. Our results are consistent with the descriptive analysis presented here.
Figure 5: Partial correlation between price and quantity, Washington (left) and other states (right)

Notes: We repeatedly estimate specification (3) across products, within retailer-state-week. We then average the results across retailers and across weeks in a quarter.

5 Model: demand and costs

We characterize the primitives of demand and costs with the goal of quantifying the effect of limited information on retailer profits. Demand for liquor is characterized by the random coefficient logit model, and estimated with standard nested-fixed point methods that incorporate micro-moments (Berry et al., 1995; ?. Petrin, 2002). We then estimate costs using the period between March 2016 and September 2016 – the last 6 months in the Washington sample.15 After reviewing the estimates, we present evidence that the model fits the data well and produces sensible implied markups. We then characterize the variation in the implied marginal costs.

15Note these are the last two quarters prior to the observed price decrease. As noted previously, we do not include the data during or after that unexplained price drop.
5.1 The demand for liquor

Consumer $i$ in market $m$ in month $t$ comes to retailer $r$ to buy groceries, and derives utility from purchasing liquor $j$:

$$u_{ijrmt} = \gamma_i + \alpha_ip_{jrm} + \delta_{jrm} + \lambda_{rt} + \xi_{jrm} + \epsilon_{ijrmt}. \quad (4)$$

In the above, $\gamma_i$ and $\alpha_i$ are household-specific intercepts and price coefficients, capturing heterogeneity in the tastes for liquor and sensitivities to prices across households. $x_{jrm}$ are observed promotion variables that vary over time.$^{16}$ $\delta_{jrm}$ are product-retailer-market fixed effects capturing tastes to different products which could differ across shoppers going to different retailers or are in different markets. $\lambda_{rt}$ are retailer-time fixed effects capturing changes over time in the demand for liquor in grocery stores, or changes in market positions for different retailers (in the liquor market). $\xi_{jrm}$ are unobserved characteristics or demand shocks. $\epsilon_{ijrmt}$ are type-1 extreme value utility shocks. If the consumer does not buy any liquor in the given trip, her utility is normalized to $u_{i0rmt} = \epsilon_{i0rmt}$.

The consumer chooses among products in a given retailer, i.e. in choice set $J_{rmt}$. This choice set assumes that liquor category does not drive store choice, and the retailers act as monopolists over their own store traffic. As noted earlier, this assumption is motivated by multiple corroborating observations including that consumers’ grocery shopping patterns after liquor deregulation are stable, and when a consumer makes a liquor purchase, she typically spends more on other grocery items than liquor on the trip, and the degree of overlap between retailer cross-shopping is limited. With these assumptions, the market share within a retailer-market is an integral of logit choice probability over random coefficients

$$s_{jrm} = \int s_{ijrmt}dF(\alpha_i) = \int \frac{\exp(\gamma_i + \alpha_i p_{jrm} + x_{jrm} \beta + \delta_{jrm} + \lambda_{rt} + \xi_{jrm})}{1 + \sum_{j' \in J_{rmt}} \exp(\gamma_i + \alpha_i p_{j'rm} + x_{j'rm} \beta + \delta_{j'rm} + \lambda_{rt} + \xi_{j'rm})}dF(\alpha_i, \gamma_i). \quad (5)$$

$^{16}$These are indicator variables of whether the product is on feature or display, or feature/display status unknown.
To capture heterogeneous tastes, we follow Conlon and Rao (2015) and parameterize the household intercept as a function of log household income,

\[ \gamma_i = \gamma_0 + \gamma_1 \log(y_i), \]  

and the price coefficient as a function of log household income and independent random draws

\[ \alpha_i = \alpha_0 + \alpha_1 \log(y_i) + \sigma_v v_i, \]  

where \( v_i \) is drawn from independent standard normal distributions.

### 5.2 Identification

#### 5.2.1 Price coefficient

We estimate model parameters by the set of moments that demand shocks \( \xi_{jrm} \) are conditional mean zero, given instruments \( z_{jrm} \) (including non-price covariates, fixed effects, and excluded instruments for price and random coefficients):

\[ \mathbb{E}[\xi_{jrm} | z_{jrm}] = 0. \]  

Despite the inclusion of fixed effects, retailers might set prices based on private information not directly controlled for by these covariates: for example, the retailer might set higher prices for certain products in markets or time periods with high demand, in which case prices will be correlated with \( \xi_{jrm} \) and the price coefficient estimate will be biased.

To address the above problem, we construct price instruments similar to Conlon and Rao (2015). For each product, we take quantity-weighted prices across all states other than Washington, and use these prices as instrument for retail prices in Washington. The prices in other states likely capture wholesale price variations that are common across states but do not correlate with demand shocks in Washington (after controlling for the above fixed effects). One example of
wholesale price co-movement is that prices of Scotch and Irish whisky move with the USD-GBP exchange rate, as illustrated in Figure 19 in the Appendix.

5.2.2 Random coefficients

We instrument the random coefficients by combining “BLP” instruments and micro moments. First, we count the number of products available in each retailer-market-month, which are typically referred to as “BLP” instruments after Berry et al. (1995). Variations in the market shares of the focal product in response to changes in the number of products identify the substitutability to other products versus to the outside option, which is captured by the random intercept $\gamma_i$.

Second, we construct two sets of micro moments using the household panel data, to help identifying the interaction terms with log income. Specifically, we divide annual household income (in thousand dollars) into four bins $I_b$: (0, 42.5], (42.5, 85], (85, 125], (125, $\infty$). The first three bins correspond to 33% and 67% quantiles for income below $125,000. The last bin is chosen as all households within this category are grouped into one bin in the Nielsen Homescan data.

Next, for each income bin, we compute the average probability of buying liquor given trips to the retailer,\(^\text{17}\) as well as the average price paid given liquor purchases. Then, for each set of parameters $\theta$ in the structural model, we compute the average probability of choosing the inside good,$\^\text{18}$

$$s^b_{j\text{rmt}} = \frac{1}{N_b} \sum_{i \in I_b} \sum_{j \in J_{\text{rmt}}} s_{ij\text{rmt}}(\theta)$$

and the average price paid given purchase of liquor:

$$p^b_{j\text{rmt}} = \frac{1}{N_b} \sum_{i \in I_b} \frac{\sum_{j \in J_{\text{rmt}}} p_{j\text{rmt}} \cdot s_{ij\text{rmt}}(\theta)}{\sum_{j \in J_{\text{rmt}}} s_{ij\text{rmt}}(\theta)}$$

where $N_b$ is the number of income draws falling into bin $b$. Finally, as in Petrin (2002), we match the observed purchase probabilities and purchase prices to the simulated ones as our micro moments.\(^\text{18}\)

\(^{17}\)To be exact, we construct this probability at the household-retailer-month level.

\(^{18}\)Conlon and Rao (2015) use similar moments but impose that these moments exactly hold using equality con-
5.3 Implementation details

Product set. We already restrict attention to the broader whisky category. To further simplify our setup, we restrict attention to liquor products with the size of 750ml, restricting to 176 products (out of 260 from grocery retailers) that take 63.6% overall grocery liquor revenue. Focusing on one size alleviates the necessity of having to model non-IIA substitution between sizes. Miravete et al. (2017) characterize substitution between different categories (e.g. Whisky or Vodka) and sizes with random coefficients on these characteristics. 19

Aggregation across stores and weeks. The original data is on the level of product-retailer-store-week. We aggregate the data to product-retailer-market-month (market defined as 3 digit zip code) for two reasons. First, liquor is a slow-moving product and there are often weeks where products have 1 or 2 unit sales. It is entirely plausible that some products have zero sales in some stores for considerable number of periods. In these cases, a random coefficient logit model is not well-defined because \( \log(s_{jrm}) = \log(0) \). Aggregating the sales will considerably alleviate the problem of zero shares. Second, while there are little price variations across stores within a market, we do average over variations in prices across weeks within a month. However, because liquor is a storable product, promotional elasticities could reflect forward purchasing rather than regular price elasticities used for setting long-run price levels (Hendel and Nevo, 2006). Using monthly data helps us to focus on long-run rather than short-run price response.

Fixed effects. In implementation, we control for all product-retailer and retailer-market fixed effects, instead of the full product-retailer-market fixed effects. Controlling for too many fixed effects will eliminate much statistical power and risk overfitting the data. Likewise, we include retailer-year fixed effects and common year-month fixed effects. In total, we have 405 product-retailer and retailer-market fixed effects and 66 retailer-year and year-month fixed effects (relative 19As a minor point, we also restrict the set of inside good to products that have sold at least 2,500 bottles in total, and have prices below $80. This selection criteria leaves us with 90 products but this set of products occupy 98.7% of the total revenue (98.4% of sample size after the previous sample reduction).
Market size definition. We measure market size in the following way. We take the population in the market,\textsuperscript{20} multiply it by the share of total grocery expenditure among the set of focal retailers in the given market-month, and multiply this by 2 to allow each person to purchase at most 2 bottles of liquor a month. With the above definition, the median market share is 0.00009. The median outside good share is 0.985 and the minimum outside good share is 0.812.

Estimation and inference. We estimate model parameters $\theta$ via iterative generalized methods of moment (GMM). We first stack all moments denoted $g(\theta) = \begin{pmatrix} g_1(\theta) \\ g_2(\theta) \end{pmatrix}$, where $g_1$ represents moments from the aggregate data

$$
\mathbb{E}[g_1(\theta)] = \mathbb{E}\left[ \hat{\xi}(\theta) \cdot z \right] = 0 
$$

(11)

and $g_2$ moments from micro data

$$
\mathbb{E}[g_2(\theta)] = \mathbb{E}\left[ \begin{array}{c} \hat{s}^b(\theta) - \bar{s}^b \\ \hat{p}^b(\theta) - \bar{p}^b \end{array} \right] = 0
$$

(12)

and $\bar{s}^b$ and $\bar{p}^b$ come from the household panel. The GMM minimizes the quadratic function of moments given the weighting matrix $W$:

$$
\mathbb{E}[g(\theta)]' \cdot W \cdot \mathbb{E}[g(\theta)].
$$

(13)

We start with the identity matrix as the initial value of $W$, and estimate $\theta$ by minimizing (13). Then, we take the previous estimate of $\hat{\theta}$ to compute $\hat{W} = \mathbb{E}\left[ g(\hat{\theta}) \cdot g(\hat{\theta})' \right]$, and use it to estimate $\theta$ and re-compute $\hat{W}$. We repeat this process until the estimates of $\theta$ does not change.

\textsuperscript{20}We linearly interpolate the each pair of annual population levels to obtain the monthly population levels.
Following Hansen (1982) and Petrin (2002), we compute the asymptotic variance-covariance matrix of the parameters,
\[
V = (\Gamma \Gamma^\prime)^{-1} \cdot (\Gamma W \Gamma^\prime) \cdot (\Gamma \Gamma^\prime)^{-1},
\]
where $\Gamma$ is the Jacobian matrix $\frac{\partial g(\theta)}{\partial \theta}$. Unlike Petrin (2002), the upper off-diagonal of the Jacobian is non-zero because aggregate moments $g_1(\theta)$ is informative of the random coefficients (due to “BLP” instruments).

### 5.4 Estimation results

Table 2 reports parameter estimates for the mean and standard deviation of price coefficients. We control for product-retailer, retailer-market, brand-retailer-market, retailer-year and year-month fixed effects but these are not directly reported in the table. We also estimate the model without household-level coefficients and without micro moments as in Berry (1994).

In the random coefficient logit model, we find considerable heterogeneity in price sensitivities and in category utility (i.e. the intercept). The 5th percentile price sensitivity is -0.353 and the 95th percentile is -0.169 – the former is more than twice as large as the latter. Some of this heterogeneity is driven by income, but a significant fraction is driven by other heterogeneity (which we capture by the normal random coefficient draw). In agreement with Conlon and Rao (2015), we also find that high-income consumers derive lower utility from the liquor category despite being less price sensitive.

We use model estimates to recover that $\xi_{jrmt}$ is auto-correlated.\(^{21}\) When imposing an AR(1) structure, i.e.
\[
\xi_{jrmt} = \rho \xi_{jrmt-1} + t_{jrmt},
\]
we find that $\rho = 0.624$ (standard error = 0.002).

\(^{21}\)An alternative is to estimate $\rho$ simultaneously with other structural parameters, as in Doraszelski et al. (2016).
<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Berry (1994)</th>
<th>price (first stage)</th>
</tr>
</thead>
<tbody>
<tr>
<td>price ($\alpha_0$)</td>
<td>-0.630</td>
<td>-0.180</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>price $\times \log$ (income) ($\alpha_1$)</td>
<td>0.085</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>std. dev. of price coef. ($\sigma_Y$)</td>
<td>0.056</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>intercept $\times \log$ (income) ($\gamma_1$)</td>
<td>-0.142</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.101)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>feature or display</td>
<td>0.126</td>
<td>0.116</td>
<td>-0.281</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.013)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>missing feature/display</td>
<td>0.187</td>
<td>0.185</td>
<td>-0.095</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.008)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>average price in other states</td>
<td></td>
<td></td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>number of products</td>
<td></td>
<td></td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>product-retailer and retailer-market FE ($\delta_{jrm}$)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>retailer-year and year-month FE ($\lambda_{rt}$)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Number of observations</td>
<td>165,105</td>
<td>165,105</td>
<td>165,105</td>
</tr>
<tr>
<td>R-squared (linear part)</td>
<td>0.860</td>
<td>0.841</td>
<td>0.972</td>
</tr>
</tbody>
</table>

Notes: This table reports parameter estimates of the demand side. The first column reports estimates and standard error of the main specification. The second column reports estimates of a Berry (1994) logit model. The third column reports the first stage for price in the Berry (1994) logit model. The F-statistics for the two excluded instruments is 248.
Table 3: Example of implied elasticities and markups

<table>
<thead>
<tr>
<th>Product</th>
<th>Price</th>
<th>% Margin</th>
<th>Elasticity of: 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.00</td>
<td>0.379</td>
<td>-2.248</td>
<td>0.006</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>2</td>
<td>12.16</td>
<td>0.305</td>
<td>0.003</td>
<td>-2.820</td>
<td>0.003</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>3</td>
<td>17.63</td>
<td>0.244</td>
<td>0.008</td>
<td>0.009</td>
<td>-3.562</td>
<td>0.013</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>4</td>
<td>21.06</td>
<td>0.223</td>
<td>0.013</td>
<td>0.017</td>
<td>0.022</td>
<td>-3.908</td>
<td>0.033</td>
<td>0.037</td>
</tr>
<tr>
<td>5</td>
<td>28.02</td>
<td>0.200</td>
<td>0.005</td>
<td>0.007</td>
<td>0.010</td>
<td>0.012</td>
<td>-4.433</td>
<td>0.019</td>
</tr>
<tr>
<td>6</td>
<td>29.91</td>
<td>0.197</td>
<td>0.001</td>
<td>0.001</td>
<td>0.002</td>
<td>0.003</td>
<td>0.003</td>
<td>-4.565</td>
</tr>
</tbody>
</table>

Notes: Elasticity and implied markup for six products (these products are picked because of the differences in prices), sold by retailer 32 in July, 2016. The elasticity table reads: 10% decrease in price of product 1 will increase its sales by 22.48% and decrease the sales of product 2 by 0.06%.

5.4.1 Implied elasticities and markups

We compute implied elasticities using our demand estimates. In table 3, we present the own- and cross- elasticity matrix for six products sold by retailer 32 in July, 2016 (close to the terminal period of the sample where prices will be assumed optimal). We find that elasticities are increasing in magnitude with price: the implied price elasticities of these example products range between -2.25 and -4.43, and are very close to Miravete et al. (2017), who find elasticities in the Pennsylvania liquor market to be -2.9 for cheap products and -4.9 for expensive products. Intuitively, this finding comes from the fact that consumers who are less sensitive to price (and thus are the main customers for high-end liquor) value liquor category lower as a whole and thus have limited willingness to pay. Thus, retailers have lower percentage margins for high-price products.

Based on those elasticities, we compute the implied margin as a share of price if retailers set prices with full information about the demand parameters (the details of which will be shown in the next section). For example, for product 3, the retailer prices it at $17.63 and 24.4% is gained as gross margin (variable profit).

5.4.2 Model fit

We measure the in-sample fit of the model using the R-squared for the mean utility projection, which is inverted from applying the Berry et al. (1995) contraction mapping on observed market shares given the nonlinear coefficients. We find that the model fits data well, explaining 86.0% of
the data variation.

Further, we examine the model fit year-by-year. If, after privatization, consumers learn about availability, prices or new products, demand would have been unstable shortly after the market opens. Because our model is simple on the time dimension, one would expect the model to fit worse in the early stage of the sample. However, as shown in Figure 6, the model explains the data well throughout the sample: R-squared on mean utility range between 0.84 and 0.87 during 2012-2015. This result implies that consumer learning might not be a first-order concern as it does not generate important variations in demand that are not accounted for by the model.

Figure 6: Model fit over time

Notes: Decomposition of variance in the predicted mean utility into observables (price and promotion), product-retailer and retailer-market FE, retailer-year and year-month FE, and the error term $\xi_{jine}$. The decomposition exercise is done year-by-year.

Finally, we also examine whether the model is able to predict brand share variations well, without flexible product-specific time trends (so the model largely relies on variations in observables to explain time-series variations in the data). With the exception of Fireball and Crown in the last
2 years, our estimated model is able to predict the evolution of brand shares over time for large brands throughout the sample, as shown in Figure 7.

## 5.5 Wholesale price (marginal costs for the retailers)

### 5.5.1 State-level uniform pricing (knowing the demand primitives)

In this section, we outline a price-setting model where the retailer sets prices for all of its products given full demand information. In Section 4, we show that retailer learning occur mainly in the sample period before 2015, suggesting that we can impose full-information optimal pricing in the latter part of the data. Our main estimates for marginal costs are based on the assumption that retailers price optimally in the last 6 months of the data. We have checked that our results are robust to marginal costs estimated using the last 3 months or 12 months of the sample.

Retailer $r$ in sets prices for its products, with the restriction that the price must be uniform for each product across all markets in the state (we omit the state subscript). When setting prices, the retailer knows demand primitives $\mathcal{D}_r = \left\{ \{ \delta_{jrm}, \lambda_{rt} \}_{j,m,t}, \alpha, \beta, \gamma \right\}$. In addition, we assume that the retailer does not know realized demand shock $\xi_{jrm}$, but takes the projected demand shock from its one-month lagged value, $\hat{\xi}_{jrm} = \hat{\rho}_1 \xi_{jrm-1}$, which we estimate after demand estimation. ²²

Denote $\tilde{s}_{jrm}$ the implied market shares when demand shock innovations are set to $\tilde{i}_{jrm} = 0$. The retailer as a multi-product and multi-market monopolist maximizes the total profit

$$
\max_{p_r} \sum_{j \in J_{rmt}} \sum_{m \in M_r} \left( (1 - f) p_{jrt} - c_{jrt} \right) \cdot \tilde{s}_{jrm} (p_{rt}) \cdot h_{rmt}.
$$

(16)

where $M_r$ is the set of markets the retailer operates in, $h_{rmt}$ is the exogenous market size of market $m$ for retailer $r$ in month $t$ (which is local population times the retailer’s share of grocery revenue share in the market), $c_{jrt}$ is the wholesale price (marginal cost) for product $j$ in $t$. $f = 0.17$ is the

---

²²Alternatively, one needs to take integrals over the vectors of $i_{rmt}$ for all $j$, which is a vector of continuous random variables of the dimension equal to the number of products in the market. Hitsch (2006) notes this is computational complex and also notes that inserting $\tilde{i}_{jrm} = 0$ is an empirically innocuous approximation. In our case, we insert the zero value only for the innovation term $i_{jrm}$, which likely results in an even smaller difference than Hitsch (2006) found.
Figure 7: Model fit by brand

Notes: Observed market shares in circles and model-predicted market shares in lines. The model-predicted shares use $\tilde{\xi}_{j,n}$ draws from its empirical distribution.
share of gross revenue levied by the state.\textsuperscript{23} We assume that

\[ c_{jrt} = \bar{c}_{jr} + \omega_{jrt}, \] (17)

i.e., marginal costs contain product-retailer-specific costs that are constant over time, plus cost shocks.

This profit maximization problem leads to the first-order condition such that for all \( j \),

\[ \sum_m (1 - f) \tilde{s}_{jrt} h_{rmt} + \sum_j' \sum_m ((1 - f) p_{j'rt} - c_{j'rt}) \frac{\partial \tilde{s}_{j'rt} (p_{rt})}{\partial p_{jrt}} h_{rmt} = 0 \] (18)

In matrix notation, and invert prices:

\[ p_{rt} = \frac{c_{rt}}{1 - f} - (\Delta_{rmt})^{-1} \sum_m \tilde{s}_{rmt} (p_{rt}) \cdot h_{rmt} \] (19)

where the \( j, j' \) th element of \( \Delta_{rmt} \) is \( \frac{\partial (\sum_m \tilde{s}_{jrt} h_{rmt})}{\partial p_{jrt}} \).

We calculate implied markup \( (\Delta_{rmt})^{-1} \tilde{s}_{rmt} \) (over “effective cost” \( c_{rt} / (1 - f) \)) based on demand estimates, and use them to compute retailer margin as \( (1 - f) (\Delta_{rmt})^{-1} \tilde{s}_{rmt} \) divided by price. Examples of these margins for six products are shown in Table 3.

5.5.2 Marginal cost estimates

We calculate implied marginal costs (wholesale price) for product-retailer pairs, \( c_{jrt} \), from the optimality conditions (18),\textsuperscript{24} which we impose on the last 6 months of the Washington sample. Figure 8 examines the distribution of estimated costs \( \bar{c}_{jr} \). We find that these costs have large dispersion across different products, but such dispersion is intuitive because of vertical differentiation in the category. The 5th percentile of average costs (among product-retailer pairs) is $3.64, whereas the 95th percentile is $31.96.

\textsuperscript{23}Excise tax and sales tax are calculated separately and not included in the list price.

\textsuperscript{24}With data aggregated to the product-retailer-month level.
As mentioned above, we check for robustness of the cost estimates by using the last 12 months or 3 months of the sample. In addition, we also compare the implied wholesale price, estimated from first-order conditions, against observed wholesale prices in the pre-privatization period for a set of products. Prior to mid-2012, the Washington State applies a fixed, 51.9% markup above the wholesale price for all products. We back out the wholesale price from published retail prices during 2010-2012. We manually match 51 high-volume products that are offered both in the pre- and post-periods, and compare the two wholesale prices between these products.

Although one should not expect the wholesale prices to stay unchanged after a change in the retail market structure, we still find the model-implied wholesale prices to be very close to those faced by the state. The mean wholesale price is $14.2 in the privatized market (median: $11.9), similar to the $12.3 in the pre-privatization era (median: $11.5). In addition, across products, the model-implied- and state- wholesale prices have a correlation coefficient of 0.98. These results provide external validity that the model does a good job recovering marginal costs for the retailers,
and also suggest that the marginal costs do not vary substantially during the sample period.

6 How do retailers learn?

In this section, we use the outputs from our structural model to provide a deeper investigation of how quickly retailers learn and investigate two key factors related to what drives that learning.

We begin by offering a more formal version of the analysis of Section 4, where we related the (unobserved) product quality to prices. Here we base this analysis on the structural demand estimates and show that prices are increasingly correlated with estimated demand intercepts. Next, we use the model to simulate a counterfactual scenario where retailers have full information about demand primitives and set prices optimally. This counterfactual provides the normative comparison by which we evaluate the potential and pace of learning. We find large gaps in initial profit levels that close over the first two years following market privatization.

Next, we provide evidence that the learning rates differ in meaningful ways between retailers and products. These differences provide insight into the drivers of both the initial state of knowledge and the pace of learning over time. Specifically, we show that retailers with prior experience in the liquor category, from pricing and selling liquor in other states, have initial prices closer to the full-information optimum compared to the local retailers, who do not have prior experience selling liquor. Whereas this initial state of knowledge provides a short-lived benefit to these multi-state retailers, we show that local retailers learn at a faster rate and, in terms of profit gap compared to full-information optimum, overtake the multi-state retailers by around the end of the first year. We further show that local retailers learn quickly especially from high-volume products, whereas multi-state retailers learn slowly especially on the high-volume products. In particular, our analysis on their pricing patterns are suggestive that these multi-state retailers are constrained by cross-state promotional policies.
6.1 Do prices capture product-retailer demand?

We re-examine the price correlations with the demand shocks using our structural estimates of the demand shocks. We examine the extent to which prices capture heterogeneity in local (WA state) tastes by calculating the correlation coefficient between price $p_{jrt}$ and the model-recovered average product-retailer fixed effect, $\bar{\delta}_{jr} = \frac{1}{M} \sum_{m} \delta_{jrm}$, for different periods of time. As retailers learn about local tastes, their prices should increasingly reflect the $\bar{\delta}_{jr}$’s. Figure 9 shows that the correlation coefficient between $p_{jrt}$ and $\bar{\delta}_{jr}$ increases over time in Washington, consistent with our descriptive evidence in Section 4.2. These increasing correlation coefficients imply that observed prices reflect the underlying differences in demand more and more over time. This again is evidence that retailers are able to learn about demand and improve the prices they set.

Figure 9: Correlation between observed price and average product-retailer FE in Washington

Notes: correlation coefficient between observed price and the model-recovered $\bar{\delta}_{jr} = \frac{1}{M} \sum_{m} \delta_{jrm}$. 
6.2 How do observed prices adjust over time compared to full-information decisions?

We use the structural model to simulate a counterfactual where retailers have full-information and set prices optimally. This counterfactual provides a comparison that acts as the model we compare actual decisions against. Such a model is appropriate because it allows us to evaluate how close to perfect-information behavior the firms’ initial behavior is, a signal of their initial knowledge position, and how long it takes the retailers to set prices that are close to the perfect information optimal prices.

We compute model-implied optimal prices from the mean marginal costs $\bar{c}_{jr}$ (that is, ignoring cost shocks $\omega_{jrt}$), from the first-order conditions defined in Section 5.5.1, for all retailers in all months in Washington.\(^{25}\) Specifically, we jointly solve for $p^*_{rt}$ as the implied full-information optimal price vector, defined as a solution of Bertrand-Nash prices

$$p^*_{rt} = \frac{c_r}{1 - f} - (\Delta_{rmt})^{-1} \sum_m \tilde{s}_{rmt}(p^*_{rt}) \cdot h_{rmt},$$

where as before, $\tilde{s}_{rmt}$ is the vector of market shares as a function of price, with $\tilde{\xi}_{jrm}t$ set to $\hat{\rho}_{jrm} \tilde{\xi}_{jrm} - 1$, the predictable part of the demand shock. Equation (20) needs to be solved as a fixed point, jointly for all products in each retailer-month. After obtaining $p^*_{rt}$, we compute the implied total profit for retailer $r$ in year-month $m$ across all products and markets:

$$\pi^*_{rt} = \sum_j \sum_m \tilde{s}_{jrm} (p^*_{rt})(1 - f) p^*_{jrt} - \bar{c}_{jr} \cdot h_{rmt}. $$

We contrast these “optimal full information profits” against the profit evaluated at the observed prices.

The left-hand side panel of Figure 10 plots the distribution of “relative absolute price gaps”.

\(^{25}\)Unlike Chintagunta et al. (2003), we do not simulate optimal local prices, since we observe that retailers set prices at the state level, see section 3.2.1.
i.e. the relative difference between the optimal and the observed prices,

\[
\text{relative absolute price gap} = \frac{|p^*_jr - p_{jr}|}{p_{jr}},
\]

across products for each quarter. We find that immediately after the privatization of liquor sales, prices are 12\% different from the optimal price at the median, while some products are priced up to 34\% away from the implied full information optimal price.\(^{26}\) As evidence of learning, over time, the gap narrows and we find that, in the median, prices are only about 2\% different from the corresponding optimal price by the first quarter of 2016. In addition, the range of price gaps are much narrower, with the upper-95\% quantile reaching only 12\%.

On the right-hand side panel, we plot the relative difference in the retailers’ total profit,

\[
\frac{\sum_r (\pi^*_r - \pi_{r})}{\sum_r \pi_{r}}.
\]

We find that in the initial quarter, the large price gaps translate into an 11\% difference in total variable profit. This difference drops very quickly to about 8\% in the second quarter after market privatization, and decreases slowly over time. Over the next three years, prices appear to be set with increasingly accurate information about demand, improving retailer profit and bringing behavior and outcomes closer and closer to the full-information counterfactual. In 2016, the difference in profit has decreased to between 1-3\% on average.\(^{27}\)

### 6.3 What are the determinants of learning and initial knowledge?

The previous section established that, compared to a full-information optimal pricing case, the actual prices are set with a meaningful loss in potential profits. In addition, the firm learns about demand over time, resulting in prices and profits that are closer to the full-information optimal. This overall pattern of learning, however, does not speak to the potential for heterogeneity in learning over time.

In this section, using the same price gap distributions and profit gap curves we depicted above,

\(^{26}\)The whiskers are 5\% and 95\% quantiles.

\(^{27}\)We note that the 1-3\% difference is a result of our ignoring the cost shocks and demonstrates that the cost shocks have a relatively small influence on profitability.
Figure 10: Differences in price and profit between observed data and the full-information optimal prices

Notes: We compute implied market shares and profit at the full-information price from the previous section, and then compute implied profit. Similarly, at observed price, we predict market shares and profit, setting demand shock innovations \( \iota = 0 \) and cost shocks \( \omega = 0 \) to maintain consistency. We then take difference between full-information profit and realized profit. As reference, monthly revenue from the stable set of products in Washington is around 35 million dollars.
we explore two key dimensions of heterogeneity that we expect could influence the initial knowledge and pace of learning. The first factor is whether the retailer has prior experience in other states that allow grocery retailers to sell liquor. Two grocery retailers and the two drug stores have such multi-state liquor operations, and in contrast, the other two grocery retailers only have operations in Washington. The second factor is the volume of sales of a product. High-volume products both provide more precise information about consumer demand (for learning from self) and are carried by more retailers and in more states (enabling learning from others). We now measures the heterogeneity in the initial pricing and in the pace of learning along these two dimensions.

6.3.1 How does experience shape the learning in a new market?

The first factor is whether the retailer already prices and sells liquor in other states that allow grocery sales of liquor. Having such experience in other states could provide benefits or costs to a retailer. The nature of these benefits depends both on the ability to transfer knowledge gained from the other states and the relevance of that information. In a model of Bayesian learning, transferring the prior experience would imply a stronger prior and, if that knowledge is relevant, a (more) accurate one. However, this initial knowledge set could also create pernicious information effects due to increased certainty (in Bayesian terms, a stronger prior). Such increased certainty could slow the rate of learning from new information and could be particularly problematic if the knowledge gained about demand in other states is not very relevant to Washington state. In addition, the multi-state liquor operation could lead to centralized functions that provide less managerial attention to any given state’s pricing problems or that place constraints because of zone pricing policies or other menu pricing constraints. In contrast, the Washington-only retailers are likely to have worse initial information about demand. Yet this weaker initial information set might be helpful to the firm, if it also means they respond more readily to demand shocks and learn more rapidly. In a Bayesian learning model, this would be reflected in a weaker prior that updates more in response to new data.

To evaluate the learning of multi-state retailers as compared to Washington-only retailers, we
split the retailers along this dimension and summarize the profit and pricing paths in Figure 11. The plots of the relative profit gap over time for the local and multi-state retailers provide a clear story. Local retailers start with poor profit performance with a 15% profit gap in the first quarter, but rapidly learn cutting the gap below 8% by the second quarter and to below 5% by the end of the second year. In contrast, multi-state retailers appear to start in a better information position with the initial profit gap below 10%. However, their rate of improvement is much slower with profit gaps close to 5% as late as the beginning of 2016. These results are consistent with multi-state retailers transferring knowledge from other states that allows them to achieve higher initial profits, but also reduces their learning rate either because of overconfidence, centralized management, or other constraints.

The price distributions in the bottom three plots of Figure 11 elaborates further on the profit results. In 2012, the local retailers have a wider price-gap distribution, but the median price gap is very similar to that of the multi-state retailers. By 2014, the local retailers significantly improve their pricing decisions and the median price gap shrink to 5%, surpassing the multi-state retailers. The same pattern holds by 2016, but the differences became even more dramatic. By 2016, almost the entire interquartile range of price gaps for the local retailers is below the median price gap of the multi-state retailers. This evidence is consistent with a “too” strong prior (overconfidence leading to a slower rate of learning) or with constraints that might hinder the ability to change prices as the learn about demand.

6.3.2 How does the volume of sales shape learning about demand?

In particular, for the local retailers we examine the relationship between the price gaps and the average sales volume of the product (averaged over the full data sample). The sales volume reflects the precision of demand signal as high-volume products are carried by more stores, more retailers, and are sold in more markets. However, sales volume might also be correlated with constraints: in a uniform-pricing world, products that are available in more markets or in more states also face more constraints.
Figure 11: Price and profit gap by local versus multi-state retailers

Notes: We group the sample into in-state retailers (two chains) and multi-state retailers (two other chains). Top panel: relative profit gap between observed and full-information prices, by local versus multi-state retailer over time. Bottom panel: relative price gap between observed and full-information prices, by local versus multi-state retailers in three cross-sections. Here, the boxplot indicates median, 25/75, and 5/95 percentiles, and the three columns focus on subsamples in the third quarter of 2012, 2014 and 2016.
Figure 12 presents the distribution of price gaps, of products with average sales volume below median (“low-volume”) and above median (“high volume”). The figure presents snapshots of these distributions for the third quarter of each year in the sample, and separately by local retailers (top panel) and multi-state retailers (bottom panel).

Focusing on the local retailers, we find that they make large pricing mistakes just when they enter the liquor business, and such mistakes are not much larger for low-volume products compared to high-volume ones: the median price gap is 13% for low-volume products and 11% for high-volume ones. One year later, they learn about demand and correct for their mistakes, notably for two types of products: First, retailers set prices more correctly for all high-volume products, with the median price gap shrinking to 6% and the 95th percentile shrinking to a little above 20%. This finding is intuitive as high-volume products provide more precise learning signals and are carried by other retailers (or in other states), enabling learning from others. Second, for low-volume products, the 95th percentile price gap shrinks from 34% to 25% in a year, despite the median does not change much. This is to say, whereas there are not much improvements in pricing of the low-volume products in the median, local retailers do correct for the largest mistakes in pricing these products. After another year or two, local retailers significantly improve the pricing of both high-volume and low-volume products, with median price within 5% of the optimal price for both product groups in 2014, and even closer to the optimal in 2015.

Furthermore, in Appendix Figure 21, we examine an alternative measure of pricing mistakes: the relative difference between price and full-information optimal price (without taking absolute value), or \( \frac{(p - p^*)}{p} \). We find that for local retailer chains, prices initially over-shoot relative to the optimum and are corrected over time. In addition, prices for high-volume products are corrected quickly in the median (but the dispersion shrink during a longer period of time), whereas low-volume products are persistently priced too high for more than a year. In summary, there is clear evidence that local retailers learn about demand, adjusting both their overall price levels and prices of individual products.

Next, for multi-state retailers, the learning patterns are drastically different. From Figure 12,
upon entering the Washington liquor market, multi-state retailers initially set prices that are closer to the full-information optimum compared to local retailers – and more so for the high-volume products. However, there is not much evidence for learning for more than two years after entering the market, as the relative price gap begin to shrink for both types of products only in 2015. We turn to more detailed depiction of the differences between price and optimal price in Figure 21, and draw two preliminary conclusions: First, multi-state retailers do not set prices that are too high relative to full-information. Likely, their prior knowledge about consumer demand elasticity suggest them to not pass through all or most taxes to consumers. Second, from Figure 21, we do observe learning for multi-state retailers, as the dispersion of the gap between price and optimal price shrink over time. However, in the first 1.5 years after market privatization, prices for high-volume products decrease relative to full-information optimal price by about 5% in the median. One plausible explanation of this pattern is that high-volume products are available in many markets or states, and such decrease in price are reminiscent of multi-state promotion policy. In summary, for multi-state retailers, there is evidence for learning about demand, but such evidence is much weaker and are suggestive of more complicated, possibly nation-wide, pricing policies.

7 Summary

In a new market, how quickly do firms learn about demand and consequently update their pricing strategies? We study the changes in retailers’ pricing policies in the Washington State liquor market, where the privatization of liquor sales leads to existing grocery chains entering the liquor market for the first time. This context is ideal to study our question because the policy change exogenously created new and (potentially) inexperienced sellers in an existing market where demand is stable over time, so that retailers’ learning to set prices can be isolated from many potential alternative stories.

Shortly after privatization, retail prices show large changes. We document two pieces of descriptive evidence showing that these price changes are consistent with retailers learning about de-
Figure 12: Implied relative price gaps by sales volume, local retailers (top) and multi-state retailers (bottom)

Notes: Percentage price gap (positive or negative) across time, retailer and product type, and the number of stores that carry the item. Median split by the number of stores. Number of stores is averaged across time for each product-retailer. Top: local retailers. Bottom: multi-state retailers.
demand. First, prices respond to lagged demand shocks and the rate of response declines over time. Second, prices are more correlated with sales as retailers’ experience in the market increases. We then estimate a structural model of demand and costs in order to simulate a counterfactual where retailers have full-information and set optimal prices based on this information. Compared to this full-information counterfactual, actual price decisions produce 11% less profits. This profit gap and the related gap between optimal and actual prices both shrink over time indicating that retailers are learning about demand. We then demonstrate that multi-state retailers have better initial information as reflected by smaller initial profit gaps. However, local retailers learn faster and surpass the multi-state retailers by the second year after deregulation. We also examine how the sales volume of products shapes the pace of learning. Thus, we shed light on not only the extent of retailer learning about demand when entering a new market, but also the drivers of the initial knowledge transfer and the pace of learning.
References

Adams, Brian and Kevin R Williams (2017), ‘Zone pricing in retail oligopoly’.


Bloom, Nicholas, Erik Brynjolfsson, Lucia Foster, Ron S Jarmin, Megha Patnaik, Itay Saporta-Eksten and John Van Reenen (2017), ‘What drives differences in management?’.


Conlon, Christopher T and Nirupama Rao (2015), ‘The price of liquor is too damn high: Alcohol taxation and market structure’.


DellaVigna, Stefano and Matthew Gentzkow (2017), ‘Uniform pricing in us retail chains’.


Hortacsu, Ali, Fernando Luco, Steven L Puller and Dongni Zhu (2017), ‘Does strategic ability affect efficiency? evidence from electricity markets.’.


Illanes, Gaston and Sarah Moshary (2017), ‘What is the marginal effect of entry? evidence from a natural experiment in liquor licensure’.

Jeon, Jihye (2017), ‘Learning and investment under demand uncertainty in container shipping’.

Miravete, Eugenio J, Katja Seim and Jeff Thurk (2017), ‘One markup to rule them all: Taxation by liquor pricing regulation’.


Seiler, Stephan, Anna Tuchman and Song Yao (2018), ‘The impact of soda taxes: Heterogeneity and welfare’.

Seo, Boyoung (2016), ‘Firm scope and the value of one-stop shopping in washington state’s deregulated liquor market’.


A Evidence on uniform pricing

We examine the degree of price dispersion for given products across stores and chains, and show that prices seem to be set on the chain level. To examine the cross-sectional aspects, we first estimate simple specifications of log prices on various levels of fixed effects (FEs):

$$\log(p_{jrm}) = \alpha_X + \varepsilon_{jrm},$$  \hspace{1cm} (23)

where, with an abuse of notation, the subscripts of $\alpha_X$ takes product level (i.e. $\alpha_j$), product-retailer level ($\alpha_{jr}$) or product-retailer-market level ($\alpha_{jrm}$) respectively. Because liquor products are naturally vertically differentiated, we do not interpret differences in the price levels across products. Instead, conditional on the estimated product FEs $\hat{\alpha}_j$, we focus on whether additional FEs at product-retailer or product-retailer-market levels help explain the “left-over” variations in log prices. Specifically, to examine the extent to which product-retailer FEs explain price variations, we calculate both the incremental amount of price variations explained by product-retailer FEs, $SSR^{jr} - SSR^{j}$, and the total variations not explained by the product FEs, $SST - SSR^{j}$.\(^{28}\) The ratio between the two,

$$\frac{SSR^{jr} - SSR^{j}}{SST - SSR^{j}},$$  \hspace{1cm} (24)

measures the amount of price variations explained by adding product-retailer FEs on top of the product FEs. We calculate the incremental fit in the same way for the model with product-retailer-market FEs.

Relative to having product FEs only, we find that adding product-retailer level FEs explains 30.9% of the “left over” price variations. This number indicates that there are large differences in the price levels across retailers for the same product. However, adding the product-retailer-market level fixed effects only captures 1.7% (percentage points) additional price variations, suggesting that the meaningful cross-sectional price variations happen at the product-retailer level.

\(^{28}\)Where $SST$ is the sum of squared of the dependent variable (as a measure of total variation), and $SSR$ is the sum of squared of the regression fit (as a measure of model-predicted variation).
Figure 13: Incremental explanatory power of co-variates to price variations

Notes: This figure reports incremental fit measures defined in (24), across log price regressions with different sets of fixed effects: product($j$)-retailer($r$), product-retailer-market($m$), and from this version, adding product-time trend, product-retailer time trend and product-retailer-market time. The benchmark regression we use to compare fit is one with only product fixed effects.

Next, we estimate a series of regressions with product-retailer-market FEs, but with linear time trends that vary at different levels. This is to say, we estimate a class of regressions

$$
\log(p_{jrt}) = \alpha_{jrm} + \beta_X \cdot t + \epsilon_{jrt},
$$

(25)

where we allow the coefficient on time, $\beta_X$, to vary at the product, product-retailer, or product-retailer-market level. Measuring incremental fit in the same way, we find that product-level time trend explains 48.7% left-over variations from the model with only product FEs, implying a 16.1% incremental fit than without the product-level trend. Product-retailer trends explain 6.2% more of the variation, while product-retailer-market trends only capture an additional 1.4%. This is to say, both price levels and price variations over time occur at the product-retailer level.
B Test for cross-chain substitution

In this section, we test whether demand for a given product substitutes between chains in the local market. Denote \( j \) as a product (pooled across sizes), \( r \) as a retail chain, \( z \) as a 5-digit zipcode, and \( z \) as a month. We estimate a linear model of log sales quantity on the availability of the product in the same chain and in other chains in the market:

\[
\log (q_{jrzt}) = \beta_1 n_{\text{store}_{jrzt}} + \beta_2 n_{\text{store}_{j,-r,z}} + \delta_{jrz} + \psi_{jt} + \epsilon_{jrzt}. \tag{26}
\]

Here, \( n_{\text{store}_{jrzt}} \) is the number of stores in chain \( r \) within the 5-digit zipcode \( z \) where product \( j \) is available, and \( n_{\text{store}_{j,-r,z}} \) is the number of stores in other chains (among the six focal chains) where \( j \) is available. Nielsen RMS data only identify store location up to 3-digit zipcode level, which is too large to measure spatial substitution. We proxy chain location using the most frequently-appearing household 5-digit zipcode among shoppers to the chain.\(^{29}\) In addition, we control for product-retailer-zipcode fixed effects and product-time fixed effects. With these controls, and also keep in mind that we focus on products that are available at the start and at the end of the sample, variations in the number of stores carrying the product comes from store entry and exit (see Figure 14) and from stores not carrying the product in a subset of weeks (likely, these are due to stockouts). In addition, we also test whether promotion in other chains in the market reduce sales in the focal chain. In a similar model as (26), we estimate the effect of share of stores on price promotion (defined as price at least 5% below the regular price), for the focal chain and for other chains.

Table 4 reports the estimation results. We find that if the product is available in one more store in the same chain, sale quantity increases by 26%. This finding is coherent with Seo (2016), in that local availability is important to demand. However, availability in one more store in other chains do not affect sales quantity in the focal chain. Similarly, our estimates on price promotion suggest that the share of stores where the product is on promotion does not affect sales quantity in the focal

\(^{29}\) An earlier version of Illanes and Moshary (2017) also use this approach.
Appendix Table 4: Sales quantity on availability and promotion of other retailers

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#stores carrying the product, own chain ($\beta_1$)</td>
<td>0.2578***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0063)</td>
<td></td>
</tr>
<tr>
<td>#stores carrying the product, other chains ($\beta_2$)</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0012)</td>
<td></td>
</tr>
<tr>
<td>#share of stores on promotion, own chain</td>
<td></td>
<td>0.0797***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0056)</td>
</tr>
<tr>
<td>#share of stores on promotion, other chain</td>
<td></td>
<td>0.0019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.0066)</td>
</tr>
<tr>
<td>product-retailer-market ($\delta_{jrm}$)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>product-year-month ($\psi_{jt}$)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Notes: The effect of product availability and promotion, in the own chain and in other chains, on sales quantity.

These results suggest that liquor sales do not substitute between chains.
C Additional tables and figures

![Graphs showing number of stores selling grocery (solid) or liquor (dots) over years 2010 to 2015.]

Figure 14: Number of stores selling grocery (solid) or liquor (dots)

Notes: Solid: number of distinct store IDs for each chain selling grocery. Dots: number of distinct store IDs selling liquor. For measures of stores selling liquor, we cross-checked with the number of license holders reported by WSLCB and find identical results. The 6 chains are chain 9 (top-left), 32 (top-right), 158, 182, 4901, 4904.
Figure 15: Total liquor sales revenue over time

Notes: Sales revenue for the focal 7 chains from the liquor category in half-year intervals. White: across all products. Blue: across core products defined in Section 3.1.

Figure 16: Price changes over time for each retailer-product (initial value = 0)

Notes: The y-axis represents the relative differences between price and initial price for a given product. Blue = final price lower than initial price. Red = final price higher than initial price.
Figure 17: Liquor expenditure in the household panel

Notes: Liquor expenditure in the household panel, across the focal 6 retailers, and including the state store (or former state store) and other players.
Figure 18: Overall changes in price, regular price, promotion depth and frequency

Notes: Top left: changes in average price over time. Top right: changes in regular price, defined as the max price for given product-retailer-market, in current week and the past 4 weeks. Bottom left: changes in promotion depth as percentage discount relative to regular price, given promotion incidence. Bottom right: frequency of promotion incidence, defined as indicator of promotion depth greater or equal to 5%.
Figure 19: Price ratio between Scotch/Irish whisky versus Bourbon/US-made whisky


Figure 20: Total profit under observed and model-implied optimal prices

Notes: The profits are sum of profit from the six grocery retailres in given half-year window.
Figure 21: Implied relative price gaps by out-of-state volume, local retailers (top) and multi-state retailers (bottom)

Notes: Percentage price gap (positive or negative) across time, local/multi-state retailer, and average volume of the product (across retailers) in other states. Median split by the number of stores. Number of stores is averaged across time for each product-retailer. Top: local retailers. Bottom: multi-state retailers.
Figure 22: Profit gap by retailer and time

Notes: Percentage profit gap between observed and full-information prices, by retailer over time.