IZA Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12376

An Analysis of Beverage Size Restrictions

Brian A. Bourquard
Steven Y. Wu

MAY 2019

I Z A Institute of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12376

An Analysis of Beverage Size Restrictions

Brian A. Bourquard
Ernst & Young

Steven Y. Wu
Purdue University and IZA

MAY 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

An Analysis of Beverage Size Restrictions*

Due to high levels of obesity, various government interventions have been proposed to curb the consumption of sugar-sweetened beverages (SSBs). The New York City "soda-ban," which proposed to limit the size of SSBs is among the most well-known and controversial. While public debates about beverage-size-restrictions tend to focus on how consumers are impacted, we use a nonlinear pricing model to show that, for all but extremely tight restrictions, consumer welfare would be unaffected by an enforceable restriction. However, sellers' profit would decline. While consumption is predicted to decline overall, the magnitude of the decline will vary by consumer segment.

JEL Classification: D82, I18, I31

Keywords: beverage size restrictions, health economics, nonlinear pricing, obesity, soda bans, sugar consumption

Corresponding author:
Steven Y. Wu
Purdue University
Krannert Building
403 West State Street West Lafayette
IN 47907
USA
E-mail: sywu@purdue.edu

* Accepted at the *American Journal of Agricultural Economics* in May 2019. We thank Raymond Florax, Joe Balagtas, three anonymous referees, and seminar participants at Purdue University for helpful comments. Financial support from USDA-NIFA grant award no. 2014-67023-21852 is gratefully acknowledged.

Various government interventions designed to reduce the consumption of sugar-sweetened beverages (SSB) have been proposed due to research linking high sugar consumption to weight gain, type-two diabetes (Schulze et al. 2004), cardiorenal disease, obesity (Johnson et al. 2007), and metabolic syndrome (Lustig et al. 2012). While SSB taxes have received considerable attention both in practice and in the academic literature, an often overlooked but controversial policy is SSB serving size restrictions. The most well known example is the New York City "soda-ban" which proposed to restrict the size of the largest SSB that can be sold to consumers. While the NYC soda ban was ultimately overturned in court, policy debates regarding sugar and soda consumption are ongoing and contentious. For example, to preempt future regulations on portion sizes, Mississippi passed Senate Bill 2687 (2013) which prevents counties and towns from enacting rules that restrict portion sizes. The governor of Mississippi signed the bill arguing that the bill would protect consumer freedom and choice. The implication is that consumer welfare would be reduced by portion size restrictions.

In this paper, we analyze the impact of a beverage size restriction on consumer welfare, retailer welfare, and consumption using a parsimonious nonlinear pricing model. Our key policy relevant finding is that, unless a size-restriction is so tight that it eliminates the ability of retailers to engage in second-degree price discrimination, consumer welfare will be unaffected by a size restriction. All policy induced inefficiencies will be borne by retailers, and by extension, manufacturers. Additionally, the size-restriction will have differential effects across consumer segments with the largest consumers of SSBs most likely to have their consumption curbed. From a policy perspective, this appears to be a desirable outcome since the policy may achieve its intended effect of reducing sugar consumption by those who tend to purchase larger portion sizes.

One qualification is that our results focus only on consumer welfare from the *consumption* of soft-drinks and do not incorporate potential health benefits. However, health benefits increase consumer welfare so our results hold despite the possibility that we *underestimate* the potential gains to consumers.[1]

Despite the tendency for opponents of beverage size restrictions to portray their opposition as an issue of consumer freedom and welfare, our analytic results suggest that arguments about consumer welfare are misplaced because the restriction will have little impact on consumer welfare for most reasonable size restrictions. Rather, a legitimate argument might be that a restriction might harm businesses selling SSBs.[2]

Accounting for second-degree price discrimination is important for accurate welfare analysis. For example, using a standard textbook demand curve may lead one to mistakenly conclude that a quantity restriction away from equilibrium will reduce consumer welfare via elevated price and reduced consumption. However, under a nonlinear pricing scheme, unless a restriction is so severe that it eliminates second-degree price discrimination, consumer welfare will be unaffected. We provide the intuition for this result in the subsequent section "An Intuitive Overview of our Model and Results."

Our result that firms suffer more than consumers makes intuitive sense since such regulations are restrictions on nonlinear pricing schemes. Thus, it might be more appropriate to view SSB size restrictions as limits on strategic pricing options of retailers rather than restrictions on consumer choice. Hence, it is not all that surprising that the soft-drink industry spent millions of dollars in public relations campaigns, lobbying, and legal challenges to the soda-restriction (Grynbaum 2014), which is consistent with our prediction that businesses have much to lose.

We also point out that our analysis was done using an economic model of an enforceable SSB-size regulation, stripped away of confounding complications. This is a fairly conventional economic approach that allows us to make a first pass at understanding the impact of a regulation without confounding the regulation with implementation or enforcement problems.[3] Thus, while we use the NYC soda ban as a motivating example, our analysis is more general than just this one proposal.

A general analysis is advantageous two reasons. First, the details of an actual regulatory proposal, like the NYC soda ban, is likely to be very ad hoc and plagued with inconsistencies. In the NYC example, businesses regulated by the NYC Department of Health and Mental Hygiene are subject to the ban. This implies that restaurants would be subject to the regulation but not convenience and grocery stores (which are regulated by the State). This creates a very uneven and confusing regulation, which is fraught with loopholes that complicate economic analysis. Second, proposed regulations often have enforcement problems in implementation. For example, in the NYC case, there is no rule to prevent a consumer from purchasing two small beverages rather than one large drink.

While our analysis is grounded in theory rather than specific statutory rules, we believe that it will serve as a more useful point of reference for future studies of food/beverage restrictions. If we only model a specific statutory proposal that is plagued by implementation/enforcement problems, then it will be difficult for other researchers studying SSB proposals to determine whether their results deviate from ours because of implementation/enforcement problems or because of the actual regulation. Thus, our approach is more generalizable and can serve as a useful benchmark for a wide array of proposed size-restrictions that differ in details but not in substance.

This paper is organized as follows. The next section provides the background and context for our study, including a review of the relevant literature on soda taxes and portion sizing. Next, we provide an in intuitive overview of our model and results before we introduce the formal model. The formal model is then introduced and the unregulated benchmark outcomes are established. Subsequently, we discuss the impact of the SSB size restriction on nonlinear pricing strategies of retailers. A comparison of regulated versus unregulated outcomes allows us to examine policy implications, which we discuss in the Policy Implications section. Finally, we wrap up the paper with the Conclusion section.

Background and context

Our study is the first that we are aware of that provides a rigorous economic foundation for understanding the economics of portion size restrictions. Prior debates about these restrictions in both the Mississippi and New York cases often involved seemingly impromptu appeals to consumer sovereignty that tend to be political in nature but lacking in economic substance. Thus, studies like ours are important for adding economic content to the debates.

Since the proposed New York city soda size-restriction was struck down by the court in 2014, SSB size-restrictions have largely disappeared from policy discussions in the United States.[4] However, given that our study shows that a size-restriction would have limited impact on consumer welfare, and given that some states are passing laws that prevent local governments from enacting SSB taxes (O'Connor and Sanger-Katz 2018), perhaps size-restrictions can become feasible alternatives to soda taxes again, particularly in political environments in which consumer welfare is of primary concern.

It is important to note that SSB restrictions may not be a perfect substitute for taxes in all situations. Size-restrictions are likely to be most relevant for restaurants, fast food chains, and/or convenience stores that offer fountain drinks or other options that are meant to be consumed in one sitting. It would be difficult for a size-restriction to be effective at venues where consumers can purchase SSBs to be stored and consumed in multiple sittings. As such, size-restrictions might not be as broadly applicable as, say, SSB taxes. Nevertheless, size restrictions may still be a viable policy instrument as government agencies are increasingly making package size regulations contingent on whether the food or beverage is meant to be consumed in a single sitting versus multiple sittings. For example, in 2016, the FDA announced new rules on nutrition labeling that distinguish between single serving containers versus larger containers that are meant to be shared or consumed over time (FDA 2014). SSB size restrictions might complement single sitting environments where the costs of implementation and enforcement will likely be relatively low.[5]

While we are unaware of any economic studies on SSB size-restrictions, there have been a significant number of academic studies on taxes in recent years. Given that this article is not about SSB taxes, we will only discuss papers that have relevance for our study. Debnam (2017) finds that a one-cent per-ounce SSB tax in Berkeley, California dramatically decreased SSB consumption. However, Debnam finds that there are complicated selection effects in the populations of SSB consumers and the voters who opted for the tax. High-type consumers are less price sensitive and less responsive to the tax, and in some cases may even consume more immediately after the implementation as a reaction against it. Debnam's work therefore is consistent with our assumption that there is a distribution of different consumer types and that consumption patterns are likely to be heterogeneous. These heterogeneous patterns can highlight some potential diffeences between the impact of a tax versus a size-restriction on consumption. Whereas the tax may have limited impact on high-types who tend to be less responsive to price, a size-restriction would directly reduce the serving size to H-types. Thus, it is possible that the size-restriction may be more effective in curbing H-type consumption in single-sitting situations.

One concern with any SSB regulation is that consumers might simply substitute into other unhealthy beverages. Grogger's (2017) study uses evidence from Mexicos soda tax, implemented in 2014, and finds that there is little evidence that consumers substituted other products for SSBs.

The literature on portion sizing is also relevant for our study. Portion sizes have been examined in some detail, and evidence exists that increases in portion sizes have coincided with the prevalence of obesity (Young and Nestle 2002). Chandon and Ordabayeva (2009) find that consumers are more sensitive to portion changes in only one dimension - height, width, and length - than all three dimensions. In other words, changes in size appear smaller to consumers when changes are across multiple dimensions. This suggests that one strategy for implementing a size-restriction is to implement it across multiple dimensions.[6]

Another question that might arise is whether a portion-size restriction will simply cause consumers to buy more units of the smaller serving to undo the size-restriction. Previous studies have shown consumers' intake of food items tend to track the portion or serving size, which is known as the "portion size effect" (Rolls et al. 2004; Almiron-Roig et al. 2015; English et al. 2014). This casts doubt on whether consumers would seek out larger servings or buy additional units in an attempt to achieve some idealized level of consumption that is independent of serving size. Vandenbroele et al. (2018) even show, using a field experiment, that placing smaller packages alongside the default size package can reduce purchases of a food item. This study is particularly interesting because it virtually eliminates the transaction cost of achieving the default serving size (per-unit prices were held constant) but volume sold still decreased.

An intuitive overview of our model and results

Before we discuss the formal model, we will provide an intuitive overview of our model and results. Our model is based on the assumption that retailers engage in second-degree price discrimination. Second degree price discrimination seems to be consistent with how SSBs are sold in single sitting venues as it is common for retailers to offer a menu of price/size options. Theoretically, menu pricing is used in environments where the seller is imperfectly informed about each buyer's WTP. The seller then uses a menu to induce each consumer to self-segment into the appropriate serving size. First-degree price discrimination can be ruled out because it presumes that the seller can perfectly identify the individual WTP of each consumer which is difficult to justify for restaurants or convenience stores. Third-degree price discrimination also seems to be less consistent with the stylized facts of SSB sales. Third-degree price discrimination does not rely on menus but instead charges consumers different prices based on observable information that is correlated with price elasticity and brand loyalty.[7]

Even casual observation at most restaurants and conveniences stores would suggest that menu pricing is the norm not the exception. Nonetheless, it is possible that some retailers engage in hybrid second- and third-degree price discrimination by conditioning the sizes/prices of the nonlinear pricing menu on observable information such as location, demographics, etc. that are correlated with elasticities. However, we do not build a hybrid model for two reasons. First, it is not clear to us how common hybrid pricing is whereas menu pricing is pervasive. Hence, the study of a hybrid model seems auxiliary rather than primary. A second-degree model is a parsimonious framework that probably does a good job of covering most, though not all situations. Second, it is well known that the welfare effects of third-degree price discrimination is ambiguous and highly dependent on very specific functional forms and parameterizations. Consequently, results generated from a hybrid model would likely be much more dependent on assumptions, functional forms, and parameterizations relative to a pure nonlinear pricing model that imposes only minimal curvature assumptions. Thus, we add the caveat that our welfare predictions likely hold for many, but not all situations. However, our model can easily be parameterized with specific functional forms to allow for comparative statics of how the optimal menu responds to changes in price elasticity or cross-price elasticity of demand. Hence, researchers interested in case-specific quantitative results can easily extend the model as needed to account for third-degree pricing effects.

We now provide an intuitive overview of our results both in the regulated and unregulated cases (with the SSB restriciton in place). Intuitively, nonlinear pricing is an incentive scheme designed to induce each type of buyer to self segment by choosing the serving size meant for that buyer. To maximize profit, the retailer would like to hold each buyer to his reservation utility. However, it is well known from the nonlinear pricing literature that high-WTP buyers will earn more than reservation utility by choosing a serving size meant for a low-WTP buyer.[8] Thus, high-WTP buyers will have to be paid excess surplus above their reservation utility to induce them to select the appropriate serving size. These excess rents are called "information rents" because buyers have private information about their WTP. In soft-drink sales, these information rents can come in the form of lower price per-ounce for buying larger serving sizes. Also, to reduce the size of these information rents, the retailer's optimal strategy is to reduce the size of the small serving to below the first-best level (for low-WTP consumers) to reduce its attractiveness to the high WTP buyers.

Now imagine that a size-restriction is exogenously imposed through government policy. The retailer now has to reduce the size of the largest serving. To maintain incentives for the buyers to self-select into the appropriate size, the retailer either has to make price adjustments and/or reduce the size of the smaller serving. However, because the small serving is already below first best, it becomes increasingly distortionary to reduce it further. On the other hand, a reduction in the price of the large serving can be done without introducing additional distortion. Thus, the optimal response is for the retailer to leave the small size alone and just lower the price of the (now smaller) large serving enough to ensure that the high-WTP buyers earn the same information rent as they did pre-policy. Similarly, low-WTP consumers receive only their reservation utility both before and after the policy restriction. Thus, consumer surplus for both types remains the same so long as the SSB restrictions is not so restrictive that it causes retailers to stop using nonlinear pricing. The welfare losses caused by the distortions created by the SSB restriction will be entirely borne by retailers.

The model

Our baseline model is a simple nonlinear pricing model in the spirit of Mussa and Rosen (1978) or Maskin and Riley (1984) where a profit-maximizing firm (e.g. beverage retailer) faces two types of consumers with utility functions $U(\theta_i, q) = \theta_i v(q_i) - t_i$ where q_i is quantity (e.g. size of the serving in ounces) and t_i is the price paid for q_i amount; i.e. t_i is the price for a serving of size q_i.[9] We let i=L denote a low-type (L-type) consumer, with WTP parameter, θ_L, who consumes relatively small amounts of the beverage. We let i=H denote a high type (H-type) consumer, with WTP parameter, θ_H, who consumes large amounts of the beverage. Thus, $\theta_H > \theta_L$.

We chose a two-type model over a continuous-type model because soft-drinks are typically sold with only a few sizes (e.g. small or large).[10] Moreover, we assume that $v'(q) > 0$ and $v''(q) < 0$ for all q to ensure single crossing; i.e. $U(\theta_H, q) - U(\theta_L, q)$ is increasing in q. If the utility function is differentiable in q, then single crossing implies that $\frac{\partial U(\theta_H,q)}{\partial q} > \frac{\partial U(\theta_L,q)}{\partial q}$. We will also assume that the principal's cost function is $c(q)$ such that $c'(q) > 0$ and $c''(q) > 0$ so that the cost function is increasing and convex.

The WTP parameter, θ_i, is assumed to be unobservable to the retailer so first-degree price discrimination is ruled out. The retailer only knows with probability β that a consumer will be of a L-type.[11] Thus, the retailer can only engage in second-degree price discrimination by creating a menu of price-size packages and letting the consumer "reveal" her type by self-selecting into her preferred package. We will henceforth refer to this as either the *screening* or *segmentation* pricing scheme.

Since both the retailer (principal) and the consumer (agent) behave strategically, we organize the resolution of the model as game with the following timeline of actions:

1. An exogenous (to retailers and consumers) serving size restriction is imposed.

2. The retailer chooses a discrete type of pricing scheme (to be discussed below).

3. The retailer chooses the prices and serving sizes, given the pricing scheme.

4. The consumer makes her purchase choice.

The resolution of the model is by backward induction. First, starting in stage 4, recall that the consumer's WTP is private information to the retailer. Hence, if the retailer offers more than one serving size, the retailer cannot explicitly exclude the consumer from any option. Instead, the retailer attempts to induce the consumer to "reveal" her type by offering the consumer a menu and letting the consumer choose an option. Ideally, the H-type would choose the large serving size (due to her higher WTP) while the L-type chooses the smaller size. However, the consumer will be strategic and will choose the size that yields the largest utility. Specifically, if the retailer naively designed a menu with two serving sizes for each type, each of which holds the types to their reservation utility, then the H-type consumer would be better off pretending to be the L-type by choosing the smaller size (see footnote 3). Consequently, in stage 3, the retailer needs to be strategic in choosing the optimal serving sizes and prices by ensuring that neither type has an incentive to choose the size meant for the other type. This can be accomplished by incorporating an *incentive compatibility* constraint. Moreover, the pricing of each size serving must be subject to a *participation constraint* which ensures that both types will purchase.

The principal's (retailer's) pricing design problem is to maximize expected profit subject to participation (PC) and incentive compatibility (IC) constraints:

$$\max_{q_L, q_H, t_L, t_H} \beta \left[t_L - c(q_L)\right] + (1-\beta)\left[t_H - c(q_H)\right] \quad s.t. \tag{1}$$

$$\theta_L v(q_L) - t_L \geq \overline{u} \quad (PC)$$

$$\theta_H v(q_H) - t_H \geq \theta_H v(q_L) - t_L \quad (IC)$$

$$q_H \geq 0 \qquad\qquad q_L \geq 0$$

where $\overline{u}$ is the utility derived from the next best alternative to consuming the beverage (reservation utility). A standard result in the nonlinear pricing literature is that only the L-type's participation constraint (PC) and the H-type's incentive compatibility constraint (IC) bind. Thus, we have omitted PC for the H-type and IC for the L-type. Substituting the binding constraints into the objective function yields

$$(2) \qquad \max_{q_L, q_H} \pi = \beta\left[\theta_L v(q_L) - \overline{u} - c(q_L)\right] + (1-\beta)\left[\theta_H v(q_H) - \overline{u} - (\theta_H - \theta_L)v(q_L) - c(q_H)\right]$$

$$q_H \geq 0 \qquad\qquad q_L \geq 0$$

The first order Kuhn-Tucker conditions are

$$(3) \qquad (1-\beta)\left[\theta_H v'(q_H) - c'(q_H)\right] \leq 0 \quad where \quad q_H \geq 0 \quad \& \quad \frac{\partial \pi}{\partial q_H} q_H = 0$$

$$(4) \quad \beta\left[\theta_L v'(q_L) - c'(q_L)\right] + (1-\beta)\left[-(\theta_H - \theta_L)v'(q_L)\right] \leq 0 \quad where \quad q_L \geq 0 \quad \& \quad \frac{\partial \pi}{\partial q_L} q_L = 0$$

Moving back to stage 2, note that the above K-T conditions form the basis for three cases of economic interest that will determine what type of discrete pricing scheme will be offered. In case i, the retailer serves both types of consumers. For example, with two-types, the retailer might offer two different serving sizes, such as a large and a small. The cups would be priced in an incentive compatible manner which would make it individually rational for the H-type to choose the large.[12] In case ii, only the H-type customer is served. For example, the retailer might offer only one size that just holds the H-type at its reservation utility. The L-type would not purchase because s/he would earn utility below his/her next best option. Finally, in case iii, the retailer offers a one-size-fits-all cup size to serve both types of consumers. *We will treat case i as the default case because retailers typically do offer multiple sizes to consumers. So we will assume that, prior to the size-restriction regulation, retailers are selling to both types through second degree price discrimination.*[13]

Finally, in stage 1, the serving-size restriction is implemented. Rather than treating the size-restriction as an endogenous policy variable within a full mechanism problem, we take an abbreviated approach of treating the restriction as an exogenous policy variable facing the retailer and consumer. Our approach is justified in our context because our fundamental question of interest is how consumer welfare (and to a lesser extent, producer welfare) is affected under varying levels of size-constraints. In that sense, our approach is no different from how SSB taxes are typically treated in the literature where the retailer and consumers face an exogenous tax.[14]

To model the size-restriction, we incorporate the constraint $q \leq \hat{q}$ where $\hat{q}$ is the maximum allowable size. The retailer's optimization problem, 2, becomes:

$$(5) \quad \max_{q_L, q_H} \pi = \beta \left[\theta_L v(q_L) - \overline{u} - c(q_L)\right] + (1-\beta) \left[\theta_H v(q_H) - \overline{u} - (\theta_H - \theta_L) v(q_L) - c(q_H)\right] \quad s.t.$$

$$0 \leq q_H \leq \hat{q} \qquad\qquad 0 \leq q_L \leq \hat{q}$$

Before proceeding with the analysis of the size restriction, we begin by establishing the unregulated benchmark, which would be akin to $q \leq \hat{q}$ not binding.

Benchmark case: no regulation

Without regulation, the retailer serves both types of consumers with a menu of two serving sizes. The K-T conditions, 3 and 4, become:

$$\theta_H v'(q_H) = c'(q_H) \tag{6}$$

$$\theta_L v'(q_L) = c'(q_L) + \frac{(1-\beta)}{\beta}\left[\theta_H - \theta_L\right] v'(q_L) \tag{7}$$

Thus, the retailer chooses serving sizes such that the H-type consumes the first-best amount whereas the L-type gets less than first best; i.e. a downward distortion.[15] This is a standard result and the intuition is that, because H-types would earn surplus above their reservation utility if they choose the L-type serving, the retailer has to pay an "information rent" to the H-type for choosing the H-type serving size. One way for the retailer to reduce this information rent is to decrease the L-type serving size. While this creates some inefficiency and hence revenue loss from serving L-types, the tradeoff is that the retailer can lower the information rent paid to the H-type.[16]

As a practical example, if we denote the solution to 6 as q_H^* and the solution to 7 as $\tilde{q}_L$, and 32 ounces is first-best for the H-type, then q_H^*=32. If 16 ounces is first-best for the L-type, then $\tilde{q}_L < 16$.[17]

The PC and IC constraints can be used to generate the serving prices.

(8) $$t_H = \theta_H v(q_H^*) - (\theta_H - \theta_L)v(\tilde{q}_L) - \overline{u}$$

(9) $$t_L = \theta_L v(\tilde{q}_L) - \overline{u}$$

Note that the prices are set such that the L-type's participation constraint is just satisfied so the L-type makes no rent. The H-type's price, however, is discounted by the information rent $(\theta_H - \theta_L)\tilde{q}_L v'(q_L)$. Finally, the prices and quantities can be substituted into the objective functions of the retailer (expected profit), and consumers (utility) to obtain value functions, which allow us to make welfare statements.

Proposition 1. *In the absence of a size-restriction regulation, the retailer's optimal nonlinear pricing strategy yields the following benchmark results:*

1. *The H-type is offered a serving size that yields the first-best level of consumption,* q_H^**, for a price,* t_H*, that provides an information rent driven quantity discount.*

2. *The L-type is offered a serving size,* $\tilde{q}_L$*, that is distorted downward relative to the L-type's first-best of* q_L^**. The price,* t_L*, is set to extract the L-type's rents.*

3. *The retailer's value function (maximized expected profit) is:* $\Pi = (1-\beta)[\theta_H v(q_H^*) - \overline{u} - c(q_H^*) - [\theta_H - \theta_L]v(\tilde{q}_L)] + \beta[\theta_L v(\tilde{q}_L) - \overline{u} - c(\tilde{q}_L)]$

4. *The H-type's value function (welfare under the optimal nonlinear pricing scheme) is* $U_H = \overline{u} + [\theta_H - \theta_L]v(\tilde{q}_L)$ *(earns information rents).*

5. *The L-type's value function is $U_L = \overline{u}$ (earns no excess surplus).*

We omit the proof as these are well-known nonlinear pricing results.

The impact of a size-restriction regulation

Having established the unregulated benchmark, we can now examine how a size-restriction would affect outcomes. In carrying out our analysis, we continue to refer to the game-theoretic timeline discussed earlier.

The primary impact of the size-restriction is twofold. First, in stage 3, the prices and sizes of the beverages offered to consumers are likely to change. Second, in stage 2, it might cause a discrete shift in the pricing strategies adopted by a retailer. For example, the retailer might switch from a strategy of offering a menu of options to a single-price strategy. Assesing both stages is important because examining only small continuous changes in price and quantity responses to the restriction, as one would do in a traditional demand model, is potentially misleading. Failure to account for discrete strategic shifts in strategy may lead to biased conclusions.

We will proceed as follows. First, we will examine how stage 3 prices and quantity respond to the introduction of a regulation *within each major discrete pricing strategy.* These price-quantity responses are needed to determine how the retailer's value function (profit) under each strategy will shift in response to the regulation. Once this is determined, we can look at the second effect, which is whether the regulation can induce the retailer to shift to a different discrete pricing strategy in stage 2.

Stage 3 policy impact: how do prices and quantities respond to a regulation within each discrete pricing strategy?

In this section, we focus on how price and quantities responds to a regulation *holding the discrete pricing strategy fixed*. The set of possible discrete pricing strategies are:

- Case ib (screening/segmentation): Sell to both types of consumers with a menu of differentiated H-type and L-type price-size options.
- Case iib: Sell only to H-types.
- Case iiib: Sell to both types using a one-size-fits-all pricing strategy.

Case ib: Sell to both types of consumers with a menu of differentiated H-type and L-type price-size options.

Referring back to problem 5, along with K-T conditions 3 and 4, note that when the constraint, $q \leq \hat{q}$ is binding, then the K-T conditions are

$$\theta_H v'(q_H) \geq c'(q_H) \quad \textit{where} \quad q_H = \hat{q} \tag{10}$$

$$\beta\left[\theta_L v'(q_L) - c'(q_L)\right] + (1-\beta)\left[-(\theta_H - \theta_L)v'(q_L)\right] \leq 0 \quad \textit{where} \quad q_L \geq 0 \quad \& \quad \frac{\partial \pi}{\partial q_L} q_L = 0 \tag{11}$$

In case ib, condition 10 is strictly positive whereas 11 holds with equality so that the optimal L-type serving size, $\tilde{q}_L$, must be such that $\tilde{q}_L$ satisfies $\theta_L v'(q_L) = c'(\tilde{q}_L) +$

$\frac{(1-\beta)}{\beta}\left[\theta_H - \theta_L\right] v'(q_L)$. But this is identical to 7, which suggests that if the retailer continues to use the segmentation strategy post-regulation, it will not change the serving size of L-types. Only the H-type serving size decreases from q_H^* to $\hat{q}$. Because $\tilde{q}_L$ remains unchanged, and q_H^* decreases to $\hat{q}$, it follows from equations 8 and 9 that t_H drops whereas t_L remains unchanged under the regulation.

As a practical example, imagine that prior to the regulation, the retailer offers serving sizes of q_H^*=32 ounces and $\tilde{q}_L$=12 ounces. A regulation might require that $\hat{q} = 20$ ounces. This would force $q_H = \hat{q}$=20, but $\tilde{q}_L$ remains unchanged at 12 ounces.

The key results are summarized in the following lemma.

Lemma 1. *Suppose that there is a size-restriction $q_H \leq \hat{q}$ such that the retailer continues to use a screening pricing strategy where $0 < q_L < q_H = \hat{q}$. Then*

1. *The H-type's serving size declines to $q_H = \hat{q} < q_H^*$ and t_H drops from $t_H = \theta_H v(q_H^*) - (\theta_H - \theta_L)v(\tilde{q}_L) - \overline{u}$ to $\hat{t}_H = \theta_H v(\hat{q}) - (\theta_H - \theta_L)v(\tilde{q}_L) - \overline{u}$,*

2. *The L-type's serving size, $\tilde{q}_L$, and price, t_L, remain unchanged.*

3. *The retailer's profit declines to: $\Pi_{ib} = \beta\left[\theta_L v(\tilde{q}_L) - c(\tilde{q}_L) - \overline{u}\right] +$*

   $(1-\beta)\left[\theta_H v(\hat{q}) - c(\hat{q}) - (\theta_H - \theta_L)v(\tilde{q}_L) - \overline{u}\right]$

4. *The H-type's welfare (utility) remains unchanged at $U_{Hib} = \overline{u} + [\theta_H - \theta_L]v(\tilde{q}_L)$ (still earns information rents).*

5. *The L-type's welfare remains unchanged at $U_{Lib} = \overline{u}$ (earns no excess surplus).*

Proofs are in the supplementary appendix online.

Intuitively, a forced reduction from q_H^* to $\hat{q}$ ostensibly puts pressure on incentive compatibility by lowering the l.h.s. of the IC constraint, $\theta_H v(q_H) - t_H \geq \theta_H v(q_L) - t_L$, via the term $\theta_H v(q_H)$ which is a function of q_H. To maintain incentive compatibility, the retailer can either lower t_H or q_L because the latter would reduce the r.h.s. via the term, $\theta_H v(q_L)$. However, $\tilde{q}$ is already at a lower than optimal level and reducing it further would increase the distortion and reduce revenue from serving L-types.

On the other hand, t_H can be reduced with no cost in efficiency. To see this, note that if we substitute the pricing function 8 into the l.h.s. of the IC constraint, we get $\theta_H v(q_H) - \theta_H v(q_H) + (\theta_H - \theta_L)v(\tilde{q}_L) + \overline{u} = (\theta_H - \theta_L)v(\tilde{q}_L) + \overline{u}$. Even if we replaced q_H^* with $\hat{q}$, we would still get the same result. In other words, any decrease on the l.h.s. of the IC constraint due to a reduction from q_H^* to $\hat{q}$ can be completely offset by an appropriate reduction in t_H.[18] Hence, the optimal response to a size-restriction is to reduce the price of the large serving rather than reduce the size of the small serving.

Case iib: Sell to only high types with $q_L = 0$

In this case, the retailer serves only H-type consumers and decides that it is too costly in terms of information rents to also serve L-types. Neither 10 nor 11 hold with strict equality so we have $q_H^* = \hat{q}$ and $\tilde{q}_L = 0$. Because the regulation causes q_H^* to drop to $\hat{q}$, it follows that the price charged to H-types also drops from $t_H^* = \theta_H v(q_H^*) - \overline{u}$ (from case ii) to $\hat{t}_H = \theta_H v(\hat{q}) - \overline{u}$. This price drop is due to a smaller serving size not because the retailer is granting a quantity discount driven by information rents.

Lemma 2. *Suppose that there is a regulatory restriction of the form $q_H \leq \hat{q}$ and the retailer* *serves only H-type consumers. Then*

1. *The H-type's serving size declines to* $q_H = \hat{q} < q_H^*$ *and* t_H *drops from* $t_H^* = \theta_H v(q_H^*) - \overline{u}$ *to* $\hat{t}_H = \theta_H v(\hat{q}) - \overline{u}$*.*

2. *The retailer's profit declines to:* $\Pi_{iib} = (1-\beta)[\theta_H v(\hat{q}) - c(\hat{q}) - \overline{u}]$

3. *The H-type's consumer welfare is:* $U_{Hiib} = \overline{u}$ *(no excess rents).*

A key point is that when the retailer only serves H-type consumers, it no longer needs to pay an information rent because it offers H-types only one price-size option and therefore need not worry about incentivizing H-types to choose the "right" option.

Case iiib: Sell to both types with a one-sized fits all package

Another option is for the retailer to use a one-size-fits-all strategy. The optimal one-size-fits-all strategy under a regulation is generated by solving:

$$\max_{t,q} [t - c(q)] \quad s.t. \tag{12}$$

$$\theta_L v(q) - t \geq \overline{u} \tag{13}$$

$$0 \leq q \leq \hat{q} \tag{14}$$

Because $\theta_L < \theta_H$, H-types will always purchase so long as L-types purchase which is why a participation constraint for H-types was not included. Solving 13 for t and substituting into the objective function yields:

$$\max_{q}] \, [\theta_L v(q) - c(q) - \overline{u}] \tag{15}$$

$$0 \leq q \leq \hat{q} \tag{16}$$

which yields the first order Kuhn-Tucker conditions:

$$\theta_L v'(q) \geq c'(q) \quad \& \quad q \leq \hat{q} \quad \& \quad \frac{\partial \pi}{\partial q}(\hat{q} - q) = 0 \tag{17}$$

Solving the K-T conditions yields the following proposition.

Lemma 3. *Suppose that there is a restriction of the form $q \leq \hat{q}$ and the retailer uses a one-size-fits-all strategy for both types of consumers. Then*

1. *The quantity offered to both types of consumers is $q = min\{q_L^*, \hat{q}\}$ where q_L^* is the first-best quantity for the L-type consumer.*
2. *The price is $t = \theta_L v(q) - \overline{u}$.*
3. *The retailer's profit is: $\Pi_{iiib} = \theta_L v(q) - c(q) - \overline{u}$.*
4. *The H-type's consumer welfare is: $U_{Hiiib} = \overline{u} + [\theta_H - \theta_L]v(\hat{q})$ (excess rents).*
5. *The L-type's consumer welfare is: $U_{Liiib} = \overline{u}$ (no excess rents).*

Relative to the baseline case, consumption for H-types drops from q_H^* to q and consumption for L-types increases from $\tilde{q}_L$ to q. Using our previous example, where L-type first-best is q_L^*=16, a regulation of $\hat{q}$=20 would result in $q = 16$. However, if the regulation were $\hat{q}$=15, then $q = 15$.

It is important to note that, even if $q_L^* < \hat{q}$ (e.g. q_L^*=16 and $\hat{q}$=20) so the restriction is not binding under the one-size-fits-all strategy, we cannot rule out the possibility that the restriction could induce the retailer to switch to this strategy because the restriction *would have been binding had the retailer stayed with the segmentation strategy.* This is the subject of the next section when we examine the impact of the size-restriction on stage 2 behavior of the retailer.

Stage 2 policy impact: how does the regulation affect retailers' choice of pricing strategy?

So far, we have investigated the impact of the size-restriction on stage 3 behavior where the retailer sets the prices and quantities, holding the discrete strategy fixed. We now investigate how the restriction affects stage 2 behavior and whether it might cause the retailer to switch from one type of discrete pricing strategy to another. The key to this analysis is to examine how variations in $\hat{q}$ affect the value functions (profits) of the retailer under the various pricing strategies. The retailer will choose the strategy that yields the highest profit.

We must pay attention to a subtle but important issue related to the stringency of the size restriction, $\hat{q}$, because increasingly tighter restrictions will limit the pricing strategies available. We partition the restriction into the following regions:

- *Region 1*: $q_L^* \leq \hat{q} < q_H^*$
- *Region 2*: $\tilde{q}_L \leq \hat{q} < q_L^*$
- *Region 3*: $\hat{q} < \tilde{q}_L$.

Recall that q_H^* is the first-best level of consumption for the H-type and would be implemented by the retailer in an unregulated market. The quantity q_L^* is the first-best level for the L-type. The quantity $\tilde{q}_L$ is the optimal L-type size offered to the L-type by the retailer under the unregulated segmentation strategy (Proposition 1).

Region 1: $q_L^* \leq \hat{q} < q_H^*$

Consider a restriction that is in-between the first-best levels for the H- and L-types. Returning to our example, suppose that it is optimal to serve H-types a 32oz beverage and L-types a 12oz beverage under a segmentation strategy. Suppose that 16oz is first best for L-type. A Region 1 restriction might be anywhere between 16oz to 32oz.

A key question is whether a restriction in this region would cause the retailer to move from the baseline of the segmentation strategy (Case i) to a sell-to-only H-types strategy (Case iib). The following proposition provides the answer.

Lemma 4. *Suppose that a retailer chooses the nonlinear pricing strategy outlined in Case i that offers different price-size packages to H-types and L-types. A regulation of the form* $q \leq \hat{q}$ *cannot cause the retailer to switch to the strategy outlined in Case iib where the retailer chooses only to serve H-types. Thus, the retailer will adopt a strategy consistent with Case ib over Case iib.*

Intuitively, the key driver of whether the retailer will switch from a segmentation strategy to an exclusive H-type strategy is the tradeoff between losing L-type profits versus not having to pay H-type information rents. This is expressed as:

$$\beta\left[\theta_L v(\tilde{q}_L) - c(\tilde{q}_L) - \overline{u}\right] \geq (1-\beta)[\theta_H - \theta_L]v(\tilde{q}_L) \tag{18}$$

But 18 is completely independent of $\hat{q}$ so a regulation in Region 1 would not induce the retailer to switch away from a segmentation strategy toward a H-type only strategy.[19]

The next question is whether the restriction would cause a retailer to switch from a segmentation strategy to a *single price* strategy that serves both types (Case iiib). Returning to the example, imagine a retailer who sells soda in 32 oz and 12 oz sizes at different prices. If an SSB restriction of 20 oz is imposed, then if the retailer continues to use a segmentation strategy, it must screen using a 20 oz to the H-type. To maintain the information rent to H-types needed to segment the market, the retailer would have to either lower the price of the 20 oz soda and/or reduce the L-type size. This reduces profit from the segmentation strategy so the retailer may switch to a single size soda of 16 oz priced to serve both types.

To assess whether the retailer will make this switch, we need to determine whether a restriction causes the inequality $\Pi_{ib} \geq \Pi_{iiib}$ (profits from lemmas 1 and 3) to be reversed. Writing out this inequality explicitly, we have:

(19)

$$\beta\left[\theta_L v(\tilde{q}_L) - c(\tilde{q}_L) - \overline{u}\right] + (1-\beta)\left[\theta_H v(\hat{q}) - (\theta_H - \theta_L)v(\tilde{q}_L) - c(\hat{q}) - \overline{u}\right] \geq \theta_L v(q_L^*) - c(q_L^*) - \overline{u}$$

If the inequality holds, profit from the segmentation strategy exceeds profit from the one-size-fits-all strategy and the retailer will stay with the segmentation strategy.

Lemma 5. *The implementation of a size restriction in the range $q_L^* \leq \hat{q} < q_H^*$ will not cause a retailer to switch from a segmentation strategy to the one-size-fits all single price, single size pricing strategy identified in Case iiib.*

To summarize, any policy restriction in Region 1 will not cause the retailer to switch away from the segmentation strategy.

Region 2: $\tilde{q}_L \leq \hat{q} < q_L^*$

In this region, the size regulation is restrictive enough that it precludes the retailer from implementing its optimal one-size-fits-all serving size of q_L^*. One would assume that this would make it even less likely that the retailer will switch from a segmentation strategy to a one-size-fits all strategy. It turns out that this intuition is correct.

Lemma 6. *The implementation of a size-restriction in the range $\tilde{q}_L \leq \hat{q} < q_L^*$ will not cause a retailer to switch away from a segmentation strategy to the one-size-fits all single price, single size pricing strategy identified in Case iiib.*

Continuing with our example where 32oz is the first-best size for H-types, 16oz is the first best level for L-types, and 12oz is the optimal size for L-types under the segmentation strategy, a Region 2 restriction would impose a maximum size between 12-16 ounces. A moderate restriction such as this would not cause a retailer to switch from a segmentation to a one-size-fits all strategy.

Region 3: $\hat{q} < \tilde{q}_L$

In Region 3, the size cap is so restrictive that it eliminates the possibility of the segmentation strategy. Continuing with our example, this would be a restriction that requires serving size to be less than 12 ounces. The first-order conditions 10 and 11 would become:

$$\theta_H v'(q_H) > c'(q_H) \quad where \quad q_H = \hat{q} \tag{20}$$

$$\beta\left[\theta_L v'(q_L) - c'(q_L)\right] + (1-\beta)\left[-(\theta_H - \theta_L)v'(q_L)\right] > 0 \quad where \qquad q_L = \hat{q}(< \tilde{q}_L) \tag{21}$$

Thus, the retailer sets $q_H = q_L = \hat{q}$ so screening is no longer optimal. The retailer can no longer engage in second degree price discrimination through serving size differentiation due to the regulation. However, the retailer must still choose whether to set the price to serve both types or to serve only H-types. A higher price will cause L-types to drop out of the market but increase profit margin from serving H-types. A price low enough to cause both types to buy would reduce profit margin but increase volume sold. To determine which is the optimal pricing strategy, we must compare retailer profit in Lemma 2 to retailer profit in Lemma 3.

Lemma 7. *The implementation of a size-restriction in the range* $\hat{q} < \tilde{q}_L$ *will cause a retailer to switch from a segmentation strategy to the one-size-fits all single price strategy that serves both types if* $\frac{[\theta_H - \theta_L]v(\hat{q})}{[\theta_H v(\hat{q}) - c(\hat{q}) - \overline{u}]} \leq \beta$*. On the other hand, if* $\frac{[\theta_H - \theta_L]v(\hat{q})}{[\theta_H v(\hat{q}) - c(\hat{q}) - \overline{u}]} > \beta$*, then the retailer will only serve H-types.*

Intuitively, a large β implies that there are more L-types in the population so that there is higher likelihood that any given customer who enters the store might be an L-type. Thus, ignoring this segment would substantially decrease volume sold which may not offset the increased margin by pricing to serve only H-types. On the other hand, if β is sufficiently small, then it might be profitable for the retailer to price higher to increase profit margins from serving only H-types. The loss in volume from ignoring L-types would have a lower impact on overall profit given that a small β implies relatively fewer L-types in the population.

Policy implications: how does the beverage size-restriction affect consumption and welfare?

Our model has generated a number of results about how a SSB restrictions would impact various strategies and outcomes through a series of lemmas. We organize these results in table 1 to provide quick reference and to enhance clarity. In this section, we will discuss the policy implications of these results.

Note that our analysis is based on a partial equilibrium model in which we introduce no obvious market failure so there is limited scope for addressing normative questions about whether the government *should* intervene based on social welfare. One would have to appeal to behavioral arguments (e.g. lack of self-control, hyperbolic discounting, etc.) to justify a role for government intervention. One can also appeal to general equilibrium effects such as, for example, how a reduction in SSB consumption might positively impact health care markets by reducing the number of unhealthy insurees. But as one of the first studies on the economic effects of beverage size restrictions, our model does provide a useful starting point for assessing some of the popular arguments one hears from proponents and opponents of the policy. This type of analysis may be just as important as classic normative analysis since policy success can often hinge on whether proponents (opponents) successfully highlight benefits (costs) while downplaying costs (benefits).[20]

The political arguments used by opponents of SSB restrictions tend to focus on how consumers will be made worse off, and to a lesser extent, on how jobs will be lost in the beverage industry. Supporters tend to focus on reduced sugar consumption with the implication that this will lead to improved health outcomes. Our model can shed light on whether the various political claims have merit.

Will the restriction reduce sweetened beverage consumption?

Proponents of SSB restrictions often cite the health literature on the role of SSBs in facilitating weight gain. In a systematic review of MEDLINE publications, Malik et al. (2006) found that the weight of the scientific evidence suggests that there is a positive link between SSB consumption and body weight. Therefore, the first obvious question is: does our model predict that the restriction will reduce SSB consumption?

Proposition 2. *An enforceable SSB size restriction will reduce beverage serving size to H-types and will only reduce serving sizes to L-types if the regulation is extremely restrictive (i.e. $\hat{q} < \tilde{q}_L$).*

Referring to table 1, one can see that the results listed in the first two rows summarize Proposition 2. While a reduction in serving size does not necessarily imply reduced consumption, the literature on portion sizing discussed earlier suggests that consumption tracks portion sizes. Thus, the restriction will likely reduce SSB consumption, particularly to H-types which is likely the targeted population.

Even in cases where enforcement may be lax, there may still be some reduction in consumption in single sitting environments. Practical barriers such as having to carry two-cups instead of one or having limited cup-holders in vehicles may constrain consumption of multiple smaller sized sodas. It is also possible that consumer expectations of what constitutes a standard serving size may adjust downward over time, just as expectations have adjusted upwards over the last several decades with the introduction of increasingly larger sizes.

How will the restriction affect consumer welfare?

A claim made by opponents of the SSB regulation is that consumers will be harmed (Grynbaum 2012; Nestle 2012). Moreover, the beverage industry has argued that the restriction would be regressive and discriminatory because it would disporportionately affect low income and minority consumers (Grynbaum 2012). This claim has some merit as Han and Powell (2013) find that low-income people tend to be heavier consumers of SSBs than high-income people. Thus, low income consumers are more likely to be H-type consumers within our model.

Proposition 3. *For light to medium size restrictions in the range* $\hat{q} \in [\tilde{q}_L, q_H^*]$ *(Regions 1 and 2), consumer welfare will be unaffected. However, for highly restrictive regulations where* $\hat{q} < \tilde{q}_L$ *(Region 3) that causes a retailer to endogenously switch from a segmentation strategy* *to a one-size strategy that serves both types, the H-type consumer's welfare will decline while the L-type consumer's welfare will remain unaffected. If instead, the retailer endogenously switches to a one-size strategy that serves only H-types, then there will be welfare reductions for both types of consumers.*

Referring to table 1, one can see that the results listed in the third and fourth rows summarize Proposition 3. Consumer welfare will largely be unaffected unless the size-restriction is extreme enough to eliminate second degree price discrimination. This also means that an SSB restriction is unlikely to be regressive unless the restriction is made unusually restrictive in low-income or minority areas, a scenario that is unlikely to be feasible politically.

Note that we did not include improved health in the consumer welfare metric. We avoided doing so because it would be difficult to pin down a precise measure of consumer welfare gains from health improvements and any such attempts would have been fraught with arbitrary assumptions. As such, it would have been easy for us to generate any conclusion we wanted by strategically choosing our assumptions. Nevertheless, our result that consumers are unlikely to suffer welfare losses from a size-restriction holds despite the fact that we likely *underestimated* benefits to consumers.

Finally, by not including health benefits, our results are robust to substitution effects where consumers switch to other unhealthy products. For example, Dubois et al. (2018) show that the health benefits from banning advertising on potato chips are likely to be mitigated by consumers switching to other junk foods. By not including health benefits in our consumer welfare metric, it is as if we have assumed that SSB health benefits are zero or have been completely offset by substitution.

How will the restriction affect retailer welfare?

Opponents of SSB size restrictions also highlight the potential for job losses and harm to small businesses in the beverage industry (e.g. Kennedy (2012)). Our model shows that these claims are legitimate.

Proposition 4. *The retailer will unambiguously suffer a loss in expected profits for binding beverage restrictions of any size.*

The sell side of the beverage industry will bear the brunt of the efficiency losses from the regulation. Given the reduction in industry profits, one can naturally expect that some small businesses will be harmed and there is potential for job losses.

Conclusion

This paper studies the economic effects of size-restrictions on sugar sweetened beverages (SSBs). The goal of this paper was not to advocate for or against SSB size restrictions but to outline the economic effects, including whether the restriction will reduce SSB consumption, and what the welfare effects are to consumers and retailers.

Our key findings are that an enforceable regulatory restriction on beverage sizing will likely reduce consumption, particularly to high consumption consumers. Surprisingly, we find that consumers will not suffer welfare losses under small to moderate restrictions. Thus, claims that consumers will be hurt are likely to be based on political rather than economic considerations. All policy induced welfare losses from consumption inefficiencies would be borne by sellers. Our study is the first that we are aware of that can provide policy makers with economic insights about the potential winners and losers from beverage size restrictions.

One concern that may arise is that there is potential for leakages in that consumers can circumvent a size-regulation by traveling outside the regulated area to purchase SSBs. This has also been a concern for the efficacy of SSB taxes. For example, after Philadelpha implemented a 1.5 cent per ounce SSB tax, sales of SSBs fell substantially within the city where the tax was in force, but rose just outside the city LaVito (2017). It appears that the leakage came from consumers driving outside the city limit and stocking up on SSBs. This concern, however, is less of an issue for policies that target single-sitting consumption that occurs in restaurants or convenience stores where it is more difficult to stock-up.

We would also like to address potential limitations of our model. First, our model is only a two-type model rather than a continuous type model that is often used in theoretical papers on nonlinear pricing. We chose a two-type model because even an n-type discrete model with $n > 2$ may be overkill for our problem. Under optimal nonlinear pricing, the number of sizes in a menu equals the number of types. So a continuous type model would result in a continuous number of sizes and an $n > 3$ model would result in $n > 3$ number of sizes. This would not be consistent with stylized observations of typical single sitting fountain drink options at most restaurants or convenient stores. Having said that, some retailers offer three sizes while others offer only two sizes. In the end, we decided to go with the two-type model because it is considerably simpler than the three-type model and the most important insights such as downward distortion to the low-type, optimal consumption by the highest type, and information rents are captured by the two-type model. The three-type model would deliver additional insights, but they tend to be peripheral at the cost of substantial complexity. For example, a well known result in the nonlinear pricing literature is that, if a third-type is added who falls between the low and high types, the third-type would receive some information rent, though not as large as the high types's while having its consumption distorted downward away from first-best, but not as much as the low-type's distortion. In other words, results for the middle type would resemble a linear combination of the high and low types. It should be obvious that our key result that consumer welfare will not be impacted by the restriction would continue to apply.

Second, while our results were derived focusing on a monopolist retailer who engages in second degree price discrimination, Stole (2007) points out that, for a large class of models, the primary impact of oligopolistic competition on nonlinear pricing is to reduce distortions (e.g. to low-type consumers), and to reduce price levels.[21] Moreover, the reduction in price levels benefits consumers leaving them with greater surplus. Thus, much like our omission of health benefits, our main result that consumer welfare is unaffected by the policy holds despite the fact that we underestimated consumer welfare by using the monopoly model.

Third, we measure consumer welfare from the consumption of SSBs based on standard utility theory. We do not integrate behavioral theories or psychological welfare losses. We felt that, as an early study, we should build a model based on classic assumptions and first principles. This then provides a useful foundation for incorporating extensions that allow for a nuanced comparison of how results might be affected by different behavioral forces. Another limitation to our study is that we have no data to test the predictions of the model. However, an applied theoretical analysis is the only feasible way to study the potential impact of a soft-drink portion size restriction at this time since portion restrictions have only been proposed but not implemented in practice. Hence, there is no available data. Nonetheless, our model can provide a theoretical framework for future empirical work.

A final limitation is that we do not evaluate a portion-size restriction that includes explicit limits on quantity of servings purchased. Thus, under the assumptions of our model, we cannot rule out the possibility that consumers might purchase additional servings under a size restriction. While the literature on portion-sizing that we discussed earlier suggests that consumers might not purchase additional servings in single-sitting environments, a targeted study may be needed to draw more definitive conclusions about the quantity effect.

Footnotes

[1]In other words, our results hold despite the fact that we stacked the deck against ourselves by not including health benefits. Another way to view this is that consumer welfare will be unaffected even if SSB restrictions yield no health benefits.

[2]We recognize that policy pandering to consumers might be an effective strategy for those who oppose the policy-restrictions on food/beverage marketing. However, our goal is not to study strategic public campaigns, which might be an interesting topic for future research, but rather to highlight the economic tradeoffs of the beverage restriction policies.

[3]For example, when economists study the impact of taxes, they typically strip away the statutory details/loopholes, and ignore tax avoidance schemes.

[4]Internationally, various forms of SSB consumption restrictions have shown up such as France banning unlimited refills of SSBs in restaurants de Freytas-Tamura (2017). While not explicitly a size restriction, the French regulation is intended to achieve a similar end result by reducing consumption of SSBs in a single-serving setting.

[5]In single sitting environments, consumers often only purchase one serving of a SSB. In the worse case scenario, a size-restriction can include accompanying enforcement rules that limit the number of servings and/or refills per-customer. Limiting quantity is not unheard of in retail environments especially during promotion periods.

[6]Our model does not capture these nuances in packaging. However, specifying a model that captures these nuances would become excessively complex and would likely reduce rather than enhance clarity. Nevertheless, our model can serve as a useful starting point for further analysis and the reader should be aware that our results can either be amplified or mitigated by the manner in which the restriction is packaged.

[7]Chandra and Lederman (2018) point out that brand loyalty, as measured by the strength by cross-elasticities of demand, is a significant factor for third-degree price discrimination in oligopolistic settings.

[8]Intuitively, the small serving is designed to provide just enough consumer surplus to the low WTP consumer such that it covers that consumer's reservation utility from not purchasing. However, high WTP

consumers have higher WTP and would therefore earn excess consumer surplus by choosing the same package as low WTP buyers.

[9]Tirole (1988) (section 3.3.3 starting page 149) points out that quality discrimination is very similar to quantity discrimination in the nonlinear tariff. In fact, by simple relabeling of variables, the quality model can be transformed to the quantity model and vice versa. Thus, at a formal level, the two models are identical and yield the same results

[10]The packaging of base units (ounces) into package sizes (cup sizes) is basically a mechanism design tool for implementing nonlinear pricing. That is, the mechanism designer (in this case, the retailer) designs packages (e.g. cup size consisting of a certain number of ounces) and charge consumers a cup price, taking into account the consumers underlying per-unit demand function. The seller creates these packages (cup sizes) at a particular package price to deliver these units to the various segments in an incentive compatible manner. In other words, the package size is simply a screening tool for the seller, nothing more. Consumers do not have some inherent preference for sizes aside from the ability of each size to deliver some number of base units at for some package price. In most textbook second degree price discrimination models, there is an underlying per-unit demand function but it is not made explicit but can easily be backed out from the consumers utility structure and first order conditions. But the solution is not typically presented in terms of demand per unit because the mechanism designer creates bulk packages of units and prices these packages to induce self-selection so the unit demand operates only in the background. As a practical example, fountain soda is often priced by the cup and there is rarely a per ounce price listed. This does not mean that consumers ignore how many ounces a cup provides nor does it imply that consumers dont have an underlying demand function for quantity.

[11]In the standard nonlinear pricing model, the probability β represents the fraction of L-types in the population of consumers. The fraction of consumers of a given type is typically treated as fixed and therefore β is exogenous. The challenge for the retailer then is to find an optimal nonlinear pricing strategy to segment the L- and H-types taking the distribution of types as fixed.

[12]The L-type would never choose the large because s/he would derive less utility than choosing his/her next best option.

[13]Cases ii and iii could be relevant in that a size-restriction can potentially cause the retailer to switch from case i to one of these cases. Thus, we still provide a treatment of case ii in the supplementary appendix online. Case iii is treated in a subsequent section **The impact of a size-restriction regulation.**

[14]Our goal is not to derive the optimal size regulation, which would require a far more elaborate model and impose assumptions about the social benefits of the health regulation and where the market failures are in health markets and outcomes. If health benefits from the size-restriction were zero, then the optimal size-restriction would trivially be no restriction at all. While interesting, this would distract us from our primary goal.

[15]Downward distortion occurs because marginal cost is inflated by the amount $\frac{(1-\beta)}{\beta}\left[\theta_H - \theta_L\right] v'(q_L)$ in equation 7. This term is positive so long as $\theta_H > \theta_L$ and $v'(q_L) > 0$ which are true by assumption.

[16]The need to maintain incentive compatibility due to unobservable types means the high type will always make some rents. This contrasts first-degree price discrimination where the retailer can perfectly identify the WTP of each customer and extract that customer's rents. Moreover, second degree price discrimination does not preclude competition. Tirole points out on page 152 that most second degree price discrimination takes place in oligopolistic markets Tirole (1988)

[17]The exact size of the L-type serving size depends on the value of specific parameters.

[18]One condition that must be satisfied, however, in order to make screening possible under $\hat{q}$, is that $\hat{t}_H \geq t_L$. This is easily satisfied, however, since $\hat{t}_H = \theta_H v(\hat{q}) - (\theta_H - \theta_L)v(\tilde{q}_L) - \overline{u} \geq \theta_L v(\tilde{q}_L) - \overline{u} = t_L$ reduces to $v(\hat{q}_H) \geq v(\tilde{q}_L)$, which implies that $\hat{q}_H \geq \tilde{q}_L$ by the assumption that $v'(q) > 0$ for all q. One might also be concerned that, with a regulation, the IC constraint for the low-type may actually be relevant, even though in the unconstrained model, only the high-type IC matters. However, one can easily show that the low-type IC is also implied by $v(\hat{q}_H) \geq v(\tilde{q}_L)$ or $\hat{q}_H \geq \tilde{q}_L$

[19]We must also account for the possibility that serving the high type with the segmentation strategy will still yield positive profits given the information rent. Thus, consider the profits from serving high types under the segmentation strategy, which is $(1-\beta)$ fraction of profits:

$$\theta_H v(\hat{q}) - c(\hat{q}) - (\theta_H - \theta_L)v(\tilde{q}_L) - \overline{u} \tag{22}$$

Consider the most restrictive ban within this region; i.e. $\hat{q} = \tilde{q}_L$. Substituting $\hat{q} = \tilde{q}_L$ into 22, rearranging and canceling terms yields $\theta_L v(\tilde{q}_L) - c(\tilde{q}_L) - \overline{u}$. Note that this is always positive so long as serving the low type yields positive profits. So in general, we don't have to worry about negative profitablity from serving high-types as long as it is profitable to serve low types.

[20]We leave it to other researchers to extend our basic model in the future to examine behavioral or general equilibrium effects.

[21]Under perfect competition, there is no scope for nonlinear pricing due to the law of one price. However, the fact that we frequently observe menu pricing at SSB retailers casts doubt on perfect competition in this market.

References

Almiron-Roig, E., Tsiountsioura, M., Lewis, H., Wu, J., Solis-Trapala, I., and Jebb, S. (2015). Large portion sizes increase bite size and eating rate in overweight women. *Physiology and Behavior*, 139:297–302.

Chandon, P. and Ordabayeva, N. (2009). Supersize in one dimension, downsize in three dimensions: Effects of spatial dimensionality on size perceptions and preferences. *Journal of Marketing Research*, 46:739–753.

Chandra, A. and Lederman, M. (2018). Revisiting the relationship between competition and price discrimination. *American Economic Journal: Microeconomics*, 10(2):190–224.

de Freytas-Tamura, K. (2017). France bans free soda refills in attack on obesity, *New York Times*, New York, NY, January 27. https://www.nytimes.com/2017/01/27/world/europe/france-soda-refill-ban.html.

Debnam, J. (2017). Selection effects and heterogeneous demand responses to the berkeley soda tax vote. *American Journal of Agricultural Economics*, 99(5):1172–1187.

Dubois, P., Griffith, R., and O'Connell, M. (2018). The effects of banning advertising in junk food markets. *The Review of Economic Studies*, 85(1):396–436.

English, L., Lasschuijt, M., and Keller, K. (2014). Mechanisms of the portion size effect. what is known and where do we go from here? *Appetite*, 88:39–49.

FDA (2014). *Analysis of Impacts for the Proposed Rules on the Nutrition Facts Label*, Silver Springs, MD. http://wayback.archive-it.org/7993/20170722031303/https://www.fda.gov/downloads/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/LabelingNutrition/UCM385669.pdf.

Grynbaum, M. (2012). Soda makers begin their push against new york ban, *The New York Times*, New York, NY, July 1. http://www.nytimes.com/2012/07/02/nyregion/in-fight-against-nyc-soda-ban-industry-focuses-on-personal-choice.html?pagewanted=all.

Grynbaum, M. (2014). New york's ban on big soda is rejected by final court, *The New York Times*, New York, NY, June 26. http://www.nytimes.com/2014/06/27/nyregion/city-loses-final-appeal-on-limiting-sales-of-large-sodas.html.

Han, E. and Powell, L. (2013). Consumption patterns of sugar-sweetened beverages in the united states. *Journal of the Academy of Nutrition and Dietetics*, 113(1):43–53.

Johnson, R. J., Segal, M. S., Sautin, Y., Nakagawa, T., Feig, D. I., Kang, D.-H., Gersch, M. S., Benner, S., and Sánchez-Lozada, L. G. (2007). Potential role of sugar (fructose) in the epidemic of hypertension, obesity and the metabolic syndrome, diabetes, kidney disease, and cardiovascular disease. *The American journal of clinical nutrition*, 86(4):899–906.

Kennedy, B. (2012). Nyc soda ban: Who is affected and how they are preparing for it, *Coca Cola Company Newsletter*, Atlanta, GA, October 30. http://www.coca-colacompany.com/stories/nyc-soda-ban-who-is-affected-and-how-are-they-preparing-for-it.

LaVito, A. (2017). Philadelphia's soda tax isn't the windfall some had hoped for, *CNBC*, Englewood Cliffs, NJ, August 22. https://www.cnbc.com/2017/08/22/philadelphias-soda-tax-isnt-the-windfall-some-had-hoped-for.html.

Lustig, R. H., Schmidt, L. A., and Brindis, C. D. (2012). Public health: The toxic truth about sugar. *Nature*, 482:27–29.

Malik, V., Schulze, M., and Hu, F. (2006). Intake of sugar-sweetened beverages and weight gain: a systematic review. *The American Journal of Clinical Nutrition*, 84(2):274–288.

Maskin, E. and Riley, J. (1984). Monopoly with incomplete information. *The RAND Journal of Economics*, 15(2):171–196.

Mussa, M. and Rosen, S. (1978). Monopoly and product quality. *Journal of Economic Theory*, 18(2):301–317.

Nestle, M. (2012). Why the 'consumer choice' argument is an automatic win for big soda, *The Atlantic*, Washington, D.C., July 2. http://http://www.theatlantic.com/health/archive/2012/07/why-the-consumer-choice-argument-is-an-automatic-win-for-big-soda/259293/.

O'Connor, A. and Sanger-Katz, M. (2018). California, of all places, has banned soda taxes. how a new industry strategy is succeeding, *New York Times*, New York, NY, June 27. https://www.nytimes.com/2018/06/27/upshot/california-banning-soda-taxes-a-new-industry-strategy-is-stunning-some-lawmakers.html.

Rolls, B., Roe, L., Kral, T., Meengs, J., and Wall, D. (2004). Increasing the portion size of a packaged snack increases energy intake in men and women. *Appetite*, 42(1):63–69.

Schulze, M. B., Manson, J. E., Ludwig, D. S., Colditz, G. A., Stampfer, M. J., Willett, W. C., and Hu, F. B. (2004). Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. *JAMA*, 292(8):927–934.

Stole, L. (2007). Price discrimination and competition. In Armstrong, M. and Porter, R., editors, *Handbook of Industrial Organization, Vol. 3*, pages 2221–2299. North-Holland.

Tirole, J. (1988). *The Theory of Industrial Organization*. MIT press.

Vandenbroele, J., Slabbinck, H., Van Kerckhove, A., and Vermeir, I. (2018). Curbing portion size effects by adding smaller portions at the point of purchase. *Food Quality and Preference*, 64:82–87.

Young, L. and Nestle, M. (2002). The contribution of expanding portion sizes to the us obesity epidemic. *American Journal of Public Health*, 92:246–249.

Table 1: **Qualitative Impact of a Size-restriction (Relative to No Regulation)**

	SIZE OF THE RESTRICTION		
	Region 1 (light) $(q_L^* \leq \hat{q} < q_H^*)$	Region 2 (medium) $(\tilde{q} \leq \hat{q} < q_L^*)$	Region 3 (heavy) $(\hat{q} < \tilde{q}_L)$
H-type serving (q_H)	Decrease	Decrease	Decrease
L-type serving (q_L)	Unchanged	Unchanged	Decrease
H-type consumer surplus	Unchanged	Unchanged	Decrease
L-type consumer surplus	Unchanged	Unchanged	Weakly decreasing[a]
Producer/seller surplus	Decrease	Decrease	Decrease
Seller's optimal pricing strategy	Unchanged	Unchanged	Switch to single-price strategy[b]

[a]See Proposition 3 for conditions under which the L-type consumer surplus will decrease.

[b]The single-price strategy can be priced to serve either H-types or both types. See Lemma 7