# Beer demand and its intricate peculiarities

*Sam MacIntyre, Maximilian Zebhauser, Sebastian Wolf*

*16 November 2018*

## Contents

# 1 Introduction

An omnipresent product which transcends time, beer has always been an important staple in the lifestyle of Americans, and a defining feature of American culture. It is a market of economic significance: in 2017, the American beer market accounted for approximately \$111.4 billion[1] of sales, about 0.5% of GDP. Many other parts of the economy depend on beer sales, particularly the food and restaurant service industry. The beer market is also highly controlled, with restrictions in place on who can purchase beer, where and when it can be sold. Understanding the factors for beer demand is thus an important goal for industry leaders and policy makers alike.

The American beer market is highly differentiated, with many different beer types and brands on offer. Yet, it is also highly concentrated, with most brands being owned by a small number of large global manufacturers. Perhaps once considered a homogenous product, beer tastes are beginning to diverge and refine. The rise of craft beers and a higher level of imported products signal these changes in the market. For example, in 2017, imported beer sales grew 3.2% to reach around 18% of total sales. Are niche segregated customer groups developing within the market or are beers still easily substituted? This is one of the key questions we aim to tackle by investigating demand dynamics when separating the import market and the domestic market.

The demand for beer as a whole is generally considered inelastic but this may not hold when we consider substitution patterns within the imported sector and the domestic sector. An ardent drinker of imported beer may happily switch between a Heineken and a Leffe, but would never jump to a Budweiser (and vice versa). In this study, we aim to test this hypothesis by estimating cross-price elasticities.

Another interesting feature of the beer market is the advertising and promotional environment. Naturally, domestic beers have a distinct advantage when it comes to advertising within The United States. It is foreseeable that imported beer manufacturers try to compensate for this disadvantage by engaging in more promotions. The effect and frequency of promotions within the imported and domestic groups is an aspect of the market which will be investigated.

The overarching objectives of the report are to characterise price elasticities of demand in the imported and domestic sectors using sequentially more sophisticated models of product substitution patterns based mostly on Berry's (1994) approach. Additionally we would like to consider the effects of certain explanatory variables included in the dataset such as promotions. Including these variables in our analyses will allow us to make inferences about their impact on demand.

# 2 Data

The dataset we are using is a small subset of Dominick's scanner dataset from the University of Chicago Booth School of Business. Our subset contains around 180,000 observations in 49 stores from 1991 to 1997. We restrict the datasets to 150,000 observations after dropping the last year of data, where several beer products drop out of the dataset. The restricted data contains information on 17 different beer products, of which 29.4% are imported beers, the rest are considered domestic or mass-market. The dataset comprises of a wide range of variables including input prices (barley, wheat, electricity), promotion status and socioeconomic data.

We observe a clear difference in the average price for imported beer (\$4.91) and domestic beer (\$3.73) which perhaps suggests that imported beers are viewed as a premium product and should be differentiated in our model formulations. Further motivation for such segregation appeared when we delved into price-quantity relations for some imported and domestic beers, as shown in Figure 1.

The price-quantity relations of the domestic brands (Miller and Budweiser) seem to imply very inelastic demand, whereas the foreign brands (Lowenbrau and Corona) show a more pronounced negative relationship between price and quantity. This initial exploration provides encouragement for our proposed lines of enquiry.

---

[1]Statistic retrieved from the Brewers Association - https://www.brewersassociation.org/statistics/national-beer-sales-production-data/

Table 1: Characteristics within the domestic and foreign beer groups

| Beer Type | Min | Mean | Max | Std.Dev. |
|-----------|------|------|-------|----------|
| Imported  | 2.87 | 4.91 | 6.96  | 0.88     |
| Domestic  | 1.55 | 3.73 | 16.05 | 0.71     |

Further descriptive statistics also illustrate that we have more price deviation in the imported beers, which could be due to differing demand dynamics or more variable promotion structures. As previously mentioned, the influence of promotions on demand composes one of our main hypotheses.

Further descriptive statistics also illustrate that we have more price deviation in the imported beers, which could be due to differing demand dynamics or more varied promotion structures. As previously mentioned, the influence of promotions on demand composes one of our main hypotheses.

There is more evidence for variation when we observe the market share within groups over time for the imported beers and domestic beers (Figure 2). The market shares for the imported beers seem more volatile which encourages us to study within group dynamics through a segmented model.

Not only do we see varying relationships between price and quantity in the different groups, there appears to be some evidence for changes in these relationships over time as shown in Figure 3. This finding incentivises the inclusion of year as an explanatory variable (by using dummy variables). This is a simplistic reduction of estimating time dependent elasticities but could still provide valuable insights.

In regards to promotions, when we calculate a high-level overview of the number of promotions for imported beers compared with domestic beers, we observe that imported beers are on promotion around 30% of the time compared with domestic beers which are on promotion 22% of the time (when aggregated over all years and stores). This adds rudimentary initial evidence for our hypothesis that imported beers benefit more from promotions than domestic beers.

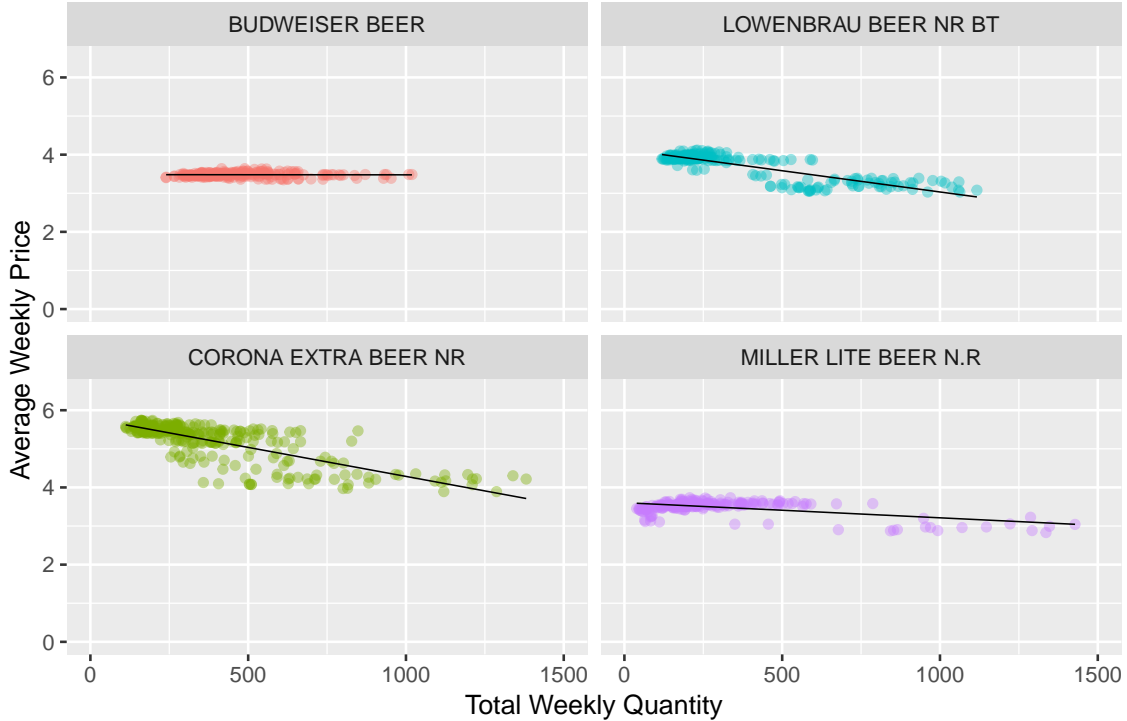Figure 1: Example price– quantity relations
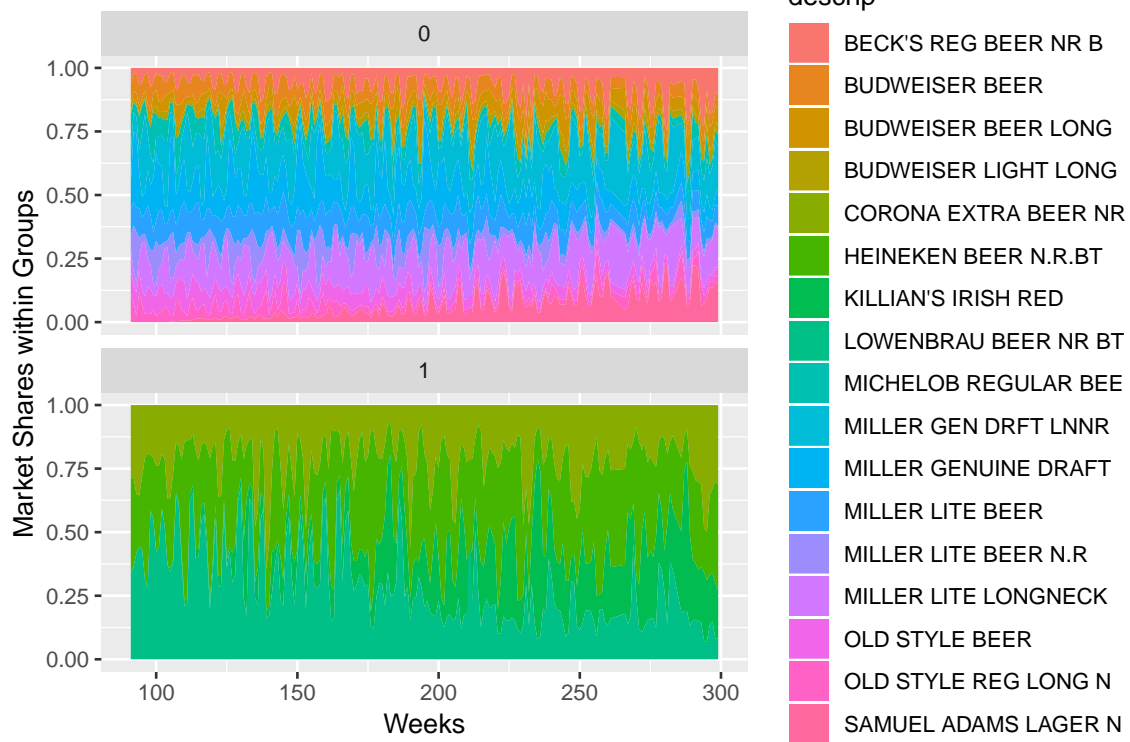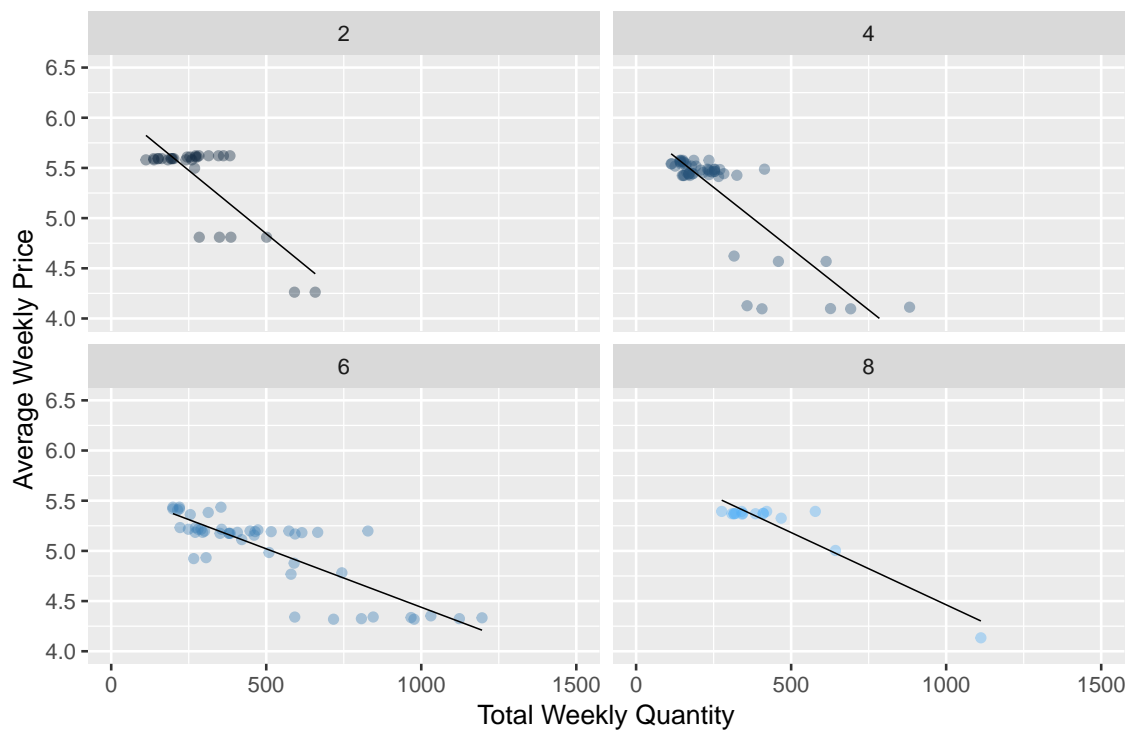
Figure 2: Market shares (by domestic/ foreign)



Figure 3: Price– quantity relations over time – Example Lowenbrau

# 3    Modelling Framework

## 3.1    Homogenous Log-Linear Demand Model

We begin by fitting a simple log-linear (multiplicative) demand model under an assumption of homogeneity across the beer market.

$$Q = AP^{-\eta}e^{\epsilon}$$

Taking this approach will allow us to infer information about the beer market as a whole and act as a natural starting point to investigate more sophisticated models. The simplicity of the multiplicative model is evident after taking logarithms of both sides of the above standard multiplicative demand equation:

$$\log(Q) = \log(A) - \eta\log(P) + \epsilon$$

As $\eta$ and $A$ now enter linearly, we can estimate these parameters using an Ordinary Least Squares approach. Note how we have included a multiplicative error term in the original model to ensure an additive error term in the log model. The error term is necessary to capture effects of omitted factors on demand. Furthermore, using a multiplicative error ensures that our model does not produce negative estimates for quantity demanded.

The simple multiplicative demand model is likely to suffer from simultaneity bias. This bias arises when the sold quantity observed is simultaenously determined by changes in the demand and the supply of the good. In other words, in a given period the price and quantity might be determined by two simultaneous shifts; consumers might shift their taste for a product, implying a shift in the demand curve, and suppliers might face increased costs, implying a shift in the supply curve. The simple multiplicative model would interpret both shifts in terms of a movement along the demand curve, and therefore mis-estimate the demand elasticity.

To mitigate this issue we use an instrumental variable that allows us to purge the price variable from its endogenous variation. The criteria for this variable are that it is correlated with the price variable, but uncorrelated or without causal relation to the quantity sold. For this purpose, the wholesale price of beer is an obvious candidate, as it is arguably closely related to the retail price, but with no or weak relation to the quantity demanded.

## 3.2    Multinomial Logit Demand Model

Assuming homogeneity across brands in the beer market is a potentially useful simplifying assumption but limits our ability to study intra-market demand dynamics. One of our goals is to investigate differences in demand and elasticities across brands, the effect of promotions, and the difference between imported and domestic beers in the US beer market. To begin investigating this problem, we can fit a Multinomial Logit Model which is derived from the customer's individual utility function, following the approach proposed by Berry (1994). This model is built on an assumed consumer utility function and allows us to derive a reduced form demand equation that is nearly as conveniently estimated as the homogenous demand model, but offers a much richer structure to exploit. It permits drawing conclusions regarding the utility consumers derive from individual product characteristics, and thus understand how companies differentiate their products from one another. In this report, we attempt to study differentiation between local and foreign brands particularly.

In the multinomial logit model we employ, each customer $i$ is assumed to derive utility from product $k$ as follows:

$$U_{ik} = x_{ik}'\beta - \alpha p_k + \xi_k + \epsilon_{ik}$$

$$U_{ik} = \delta_{ik} + \epsilon_{ik}$$

where $x'$ represents the product characteristics, $p$ the price, $\xi_k$ the unobserved product characteristics and $\epsilon$ is the error term which is log-Weibull distributed (Gumbel). $\delta_{ik}$ represents the 'mean utility'. The observable product characteristics represent all the things that we are able to measure and extract from the dataset whereas $\xi_k$ are unobservable product characteristics. That is to say that it captures information about the product that could have an influence on demand but are not measurable (or not easily measurable) such as

label design or intra-family adherence, or, in our case, whether the beer is imported or domestic. From Berry, we also obtain that the market share of product $k$ in the multinomial logit model can be represented as $s_k$ and the market share of the outside good (consumer's expenditure on other things than beer) as $s_0$.

$$s_k = \frac{e^{\delta_k}}{\sum_{j=0}^{K} e^{\delta_j}} \quad and \quad s_0 = \frac{1}{\sum_{j=0}^{K} e^{\delta_j}}$$

With the market shares in this format, model estimation is difficult, due to the non-linearity. Berry addresses this by taking log differences which ensures the important parameters enter the model linearly. From the above relations:

$$\ln(s_k) - \ln(s_0) = \delta_k$$
$$= x_k^{'} \beta - \alpha p_k + \xi_k$$

Although this approach removes the non-linearity and makes estimation considerably easier, the problem remains that prices are endogenous, since they are clearly correlated with the unobserved product character-istics or brand valuations $\xi_k$. In order to overcome this issue, we can re-employ the price instrument from the homogenous demand model; the wholesale price. To improve the suitability of this instrument, we first perform a hedonic regression of the wholesale price on the price of various inputs to purge the wholesale of any potential endogenous variation that might arise from feedback consumer choices might have on manufacturer's price-setting. Using the fitted values of the wholesale prices from the regression has the effect of cleaning up the variable to act more appropriately as an instrument.

After estimating the above reduced form regression, we can elicit elasticities as follows:

$$\eta_{jk} = \frac{\partial s_j}{\partial p_k} \frac{p_j}{s_j} = \begin{cases} -\alpha p_j (1 - s_j) & j = k \\ \alpha p_k s_k & \text{otherwise} \end{cases}$$

Note that under this model we can only recover a vector of cross-price elasticities. On the one hand, this might be considered a limitation and a strong assumption regarding market dynamics. On the other hand it significantly reduces the cost (in terms of precision) of the cross-price elasticity estimation compared to the homogenous demand model, where every cross-price had to be included as a seperate regressor. Particularly in highly differentiated markets with a large number of products, this is a strong advantage.

## 3.3    Nested Logit Demand Model

Despite taking into account heterogenous products in a single market, the multinomial logit is grounded in the strong assumption of Independence of Irrelevant Alternatives, and thus only allows relatively simple inferences to be made. A slightly more complex modelling approach invokes the Nested Logit Model. By nesting the products to further differentiate product groups (within a market), we can relax the Independence of Irrevelant Assumption (IIA) between these clusters (e.g. in our case imported and domestic beers). This is an attractive feature in our case as it allows us to explore our analysis regarding intra- and inter-group dynamics of domestic and imported beers. To implement the model, we segment the market into $G + 1$ exhaustive groups (2 groups in our case) and label them $g = 0, 1, ..., G$ and we consider the outside good as $j = 0$ as the only member of group 0. Under this model we have the utility for customer $i$ choosing product $j$ as:

$$u_{ij} = \vec{X_j}\vec{\beta} - \alpha p_j + \xi_j + \xi_{ig} + (1 - \sigma)\epsilon_{ij}$$
$$= \delta_j + \xi_{ig} + (1 - \sigma)\epsilon_{ij}$$

Note the addition of the $\xi_{ig}$ to account for the unobserved product characteristics common within groups and $\epsilon_{ij}$, which we assume are independently and identically Extreme Value Distributed errors. $\sigma$ can be interpreted as a measure of within group correlation of utilities. As $\sigma$ approaches 1, the within group correlation also

approaches 1. Using similar arguments as above regarding the market share and taking logarithms, we obtain the following log-difference equation for parameter estimation:

$$\ln(s_j) - \ln(s_0) = \delta_j + \sigma \ln(s_{j|g})$$

$$= \vec{X}_j \vec{\beta} - \alpha p_j + \sigma \ln(s_{j|g}) + \xi_j.$$

Where $s_{j|g}$ is the market share of product $j$ within group $g$. The own price and cross (within and across) price elasticities can be estimated in a similar vein to the Multinomial Logit Model.

# 4 Estimation

## 4.1 Price Instrument

As noted in the previous section, prices are likely to be endogenous in any of the demand models we want to estimate, given that the quantity of sales observed is a function both of supply shifts and demand shifts. Not accounting for this would bias our estimates. Before we proceed with our model estimations, we therefore exploit a unique feature of this dataset and create an instrumental variable using wholesale prices. Wholesale prices are expected to be exogenous to the demand decision - customers do not care about wholesale prices in their purchase decision, and manufacturers of beer are not usually expected to change the wholesale price in response to demand shifts. This of course, is a somewhat questionble assumption, and therefore we try to further improve the suitability of the instrument using a hedonic regression of the wholseale price of beer on a number of explanatory variables (barley, months, year, wages, packaging prices, electricity, store, brand). The hedonic regression is expected to remove any endogenous variation in the wholesale price, by relating it to variables that are even less likely to be impacted by shifts in the demand of beer.

We then use the fitted values of the hedonic regression to create an instrumented retail price variable by regressing the hedonic wholesale price on the retail price, and saving its fitted values for our demand estimation. The adjusted $R^2$ result of 0.813 (shown in Table 2) and a coefficient significant at the 1% level suggests that our instrument is relevant. We further confirm this with the Wald test, rendering a very large F-value, allowing us to reject the null hypothesis that the instrument is irrelevant.

In Figure 4 we plot the average instrumented retail price (grey lines) and the average real retail price (coloured lines) for four brands over time. We note that our price instrument smoothens out the effect of promotions which cause large drops and volatility for the real retail price. This is an acceptable abstraction, because we account for promotions seperately in our models. However, it also hints to the potential endogenity of promotions, which was beyond the scope of this study to explore, and for which valid instruments were difficult to construct from the available data.

Consequently, the resulting fitted retail prices based on these regressions were subsequently used as the instrumental variable in all of our further models.

Table 2: First stage regression of retail prices on (instrumented) wholesale prices

| | Dependent variable: |
| --- | --- |
| | price_weekly |
| IV_wh_price | 0.992*** |
| | (0.001) |
| Constant | 0.724*** |
| | (0.004) |
| Observations | 148,880 |
| Adjusted R$^2$ | 0.813 |

Figure 4: Real vs instrumented price
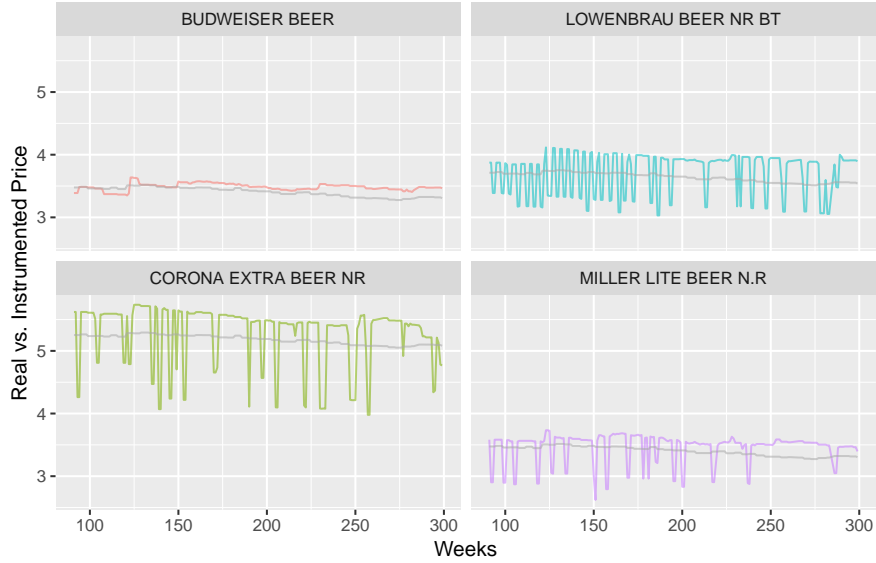


## 4.2 Estimating the Homogenous Log-Linear demand model

We begin our estimation by fitting four versions of the homogenous demand model. Firstly, by considering all the retail prices in all stores individually (1), then averaging prices and summing quantities over all stores for each brand (2), then using instrumented retail prices (3), and finally by averaging instrumented prices and summing quantities over all stores for each brand (4). The homogeneous demand model lays the foundation for more complex analyses of product differentiation to follow, by giving us an idea of the size of the demand elasticty for beer as whole, simplifying away from substitution patterns for now. Among the homogeneous demand models we estimate (4) performs best in terms of model fit, though all are relatively poor. The estimated own-price elasticity of demand falls into the range we expected (between 0 and -1) once we instrument for prices; model (3) produces -0.454 and model (4) produces -0.737. We note that averaging by product seems to improve model fit, which encourages us to include product level dummies in the models to follow.

Table 3: Results for the homogenous log-linear demand models

| | Dependent variable: | | | |
|---|---|---|---|---|
| | Log(Quantity) | | | |
| | Basic | Product totals | IV | IV, Product totals |
| | (1) | (2) | (3) | (4) |
| Log(Prices) | −1.368*** | −1.624*** | | |
| | (0.013) | (0.061) | | |
| log(IV_price) | | | −0.454*** | −0.737*** |
| | | | (0.015) | (0.074) |
| Constant | 3.808*** | 8.152*** | 2.584*** | 6.949*** |
| | (0.018) | (0.084) | (0.020) | (0.101) |
| Observations | 148,880 | 3,420 | 148,880 | 3,420 |
| Adjusted $R^2$ | 0.070 | 0.170 | 0.006 | 0.028 |

## 4.3 Estimating the Multinomial Logit and the Nested Logit Demand Models

Naturally, assuming beer is a homogenous product is invalid in most cases, given the high number of differentiated beer products on the market, and the relevance of highly differentiated products such as imported and craft beer in terms of market shares. The poor fit of the homogenous demand model may be testimony to its inadequacy. We now proceed to estimate the Multinomial Logit Model, which allows heterogeneity amongst different beers and provides us with good measures of cross-price elasticities in differentiated product markets. Besides the instrumented price, we decided to include promotional and advertising dummies, as well as store, product, month and year dummies as additional explanatory variables. We do not have data on individual product characteristics, and thus consider product dummies to fully account for differences between products. Promotional and advertising dummies indicate whether a given quantity data point was observed in a period of promotion. We differentiate between general promotions, coupon promotions, and general advertising campaigns being active. These are important variables to explore the plausibilty of our hypothesis that promotions and advertising are more important for the less-known brands, particularly foreign brands. To explore this hypothesis we include an interaction between our promotion and advertising dummies with a dummy for foreign and domestic brands. As noted earlier, a caveat of our models is that we do not have an instrument for promotions, which we would normally also expect to be endogenous.

The construction of our dependent variable has been described in the Modelling Framework section. The Nested Logit Model we estimate includes the same regressors as the Multinomial Logit Model, but adds the within-group share as an additional regressor. We create two groups, imported/foreign and domestic beers, to test whether the substitution between these groups is signifcantly different from the subsitution within these groups, or among all brands as purported by the Multinomial Logit Model. The within group market share is endogenous, and requires an instrument to allow us to estimate consistent cross-group and within group demand elasticities. We used the number of beer products sold within a particular store, the volume of overall products sold within a particular store, the number of customers that visited the store, and whether it is in an urban area to instrument for the within group-share. This assumes that the more products within a group are on being sold in a given store, the smaller an indvidual products' within-group share we expect to observe for that store. The other variables mentioned helped further improve the first-stage estimation; in our view without calling exogeneity into question.

The estimation results are presented in Table 4. In terms of explanatory power, both the Multinomial (1) and Nested Logit Model (2) appear to perform similarly well, with R2 values around 0.4. The estimated

coefficient for price effects remains at a similar level across the two models, which increases our confidence that the models are somewhat robust. We proceed to study the implied elasticities in Table 5 before returning to interpret the estimated coefficients in the context of our hypothesis.

Table 4: Results for the multinomial and nested logit demand models

|  | *Dependent variable:* | |
|---|---|---|
|  | Utility | |
|  | (1) | (2) |
| gen_promo | 1.458*** | 1.453*** |
|  | (0.056) | (0.056) |
| gen_promo_foreign | −0.156** | −0.149* |
|  | (0.077) | (0.077) |
| coupon_promo | 1.013*** | 1.008*** |
|  | (0.006) | (0.006) |
| coupon_promo_foreign | −0.139*** | −0.136*** |
|  | (0.012) | (0.012) |
| advertising_chi | 0.071*** | 0.079*** |
|  | (0.011) | (0.011) |
| advertising_chi_foreign | −0.225*** | −0.217*** |
|  | (0.011) | (0.011) |
| IV_price | −0.747*** | −0.708*** |
|  | (0.192) | (0.191) |
| IV_within_group |  | −0.301*** |
|  |  | (0.013) |
| Constant | −7.728*** | −3.668*** |
|  | (1.133) | (1.111) |
| Observations | 148,880 | 148,880 |
| Adjusted R$^2$ | 0.401 | 0.403 |

The elasticities implied by our two Logit Demand Models are presented in Table 5, and show relatively large own-price elasticity and much smaller cross-price elasticities. The size of the own-price elasticities is much larger than what we expected from the homogenous demand equation, and raises some suspicion regarding the specification of our model. Yet, in general the estimated elasticitices point in the right directions. Then Nested Logit Model further allows us to differentiate between within and across group substituion. The elasticities are notably larger for within group substitution, which lends support to our hypothesis that substitution within groups is significantly different from substitution across groups. The underlying coefficient for the instrumented price and group variable are both significant.

Table 5: Estimated elasticities - Multinomial Logit and Nested Logit Demand Models

| Brand | Own.ML | Cross.ML | Own.NL | Within.Group | Across.Group |
|---|---|---|---|---|---|
| Beck's | -4.027 | 0.0005146 | -3.001 | 0.06869 | 0.0004876 |
| Budweiser | -2.547 | 0.0002832 | -1.892 | 0.03775 | 0.0002683 |
| Budweiser Long | -2.564 | 0.0002631 | -1.9 | 0.03349 | 0.0002493 |
| Budweiser Light Long | -2.615 | 0.0001503 | -1.924 | 0.01999 | 0.0001424 |
| Corona | -3.871 | 0.0002842 | -3.032 | 0.213 | 0.0002693 |
| Heineken | -4.1 | 0.0004425 | -3.281 | 0.2958 | 0.0004192 |
| Kilians Irish | -3.498 | 0.0002921 | -2.742 | 0.1944 | 0.0002768 |
| Löwenbräu | -2.723 | 0.0002548 | -2.168 | 0.1851 | 0.0002414 |
| Michelob | -2.728 | 0.000258 | -2.019 | 0.03304 | 0.0002445 |
| Miller LNNR | -2.579 | 0.0007266 | -1.966 | 0.0883 | 0.0006885 |
| Miller | -2.563 | 0.0003125 | -1.907 | 0.04078 | 0.0002961 |
| Miller Lite | -2.572 | 0.0004128 | -1.931 | 0.05821 | 0.0003911 |
| Miller Lite N.R | -2.552 | 0.0003213 | -1.897 | 0.03884 | 0.0003044 |
| Miller Lite Longneck | -2.578 | 0.0005736 | -1.948 | 0.07008 | 0.0005435 |
| Old Style Beer | -2.546 | 0.000255 | -1.889 | 0.03529 | 0.0002416 |
| Old Style Long | -2.561 | 0.0001758 | -1.887 | 0.02242 | 0.0001665 |
| Samuel Aadams | -3.588 | 0.0004641 | -2.679 | 0.06639 | 0.0004397 |

## 4.4 Inference on hypothesis

Hypothesis 1: Are promotions working? The two promotion related parameters are positive with a high level of significance, supporting the hypothesis that promotions have a positive impact on demand. Both the Nested Logit Model and Multinomial Logit model tend to agree when it comes to direction of effects from the included variables. Our models suggest that the effect of general price promotions is about 45% higher than the effect of coupon promotions. Advertising promotions seem to have a significantly lower effect. Unfortunately, our models do not allow us to draw conlusions regarding the cost-benefit ratio of promotions, given that we do not observe the cost of a promotion in the dataset, only whether a promotion was active or not.

Hyopthesis 2: Is substitution larger within groups than across groups? Examining the out-of-group cross price elasticities from the Nested Logit Model, we find they are very close to zero. This confirms our hypothesis that consumers are unlikely to subsitute with beers outside their own preferred market segment (imported or domestic). Cross-referencing the within group elasticities with the beer brands, it is clear that the absolute values of the cross-price elasticities of the imported beers are larger. This validates our original hypothesis that consumers more readily substitute between imported beers than between imported and domestic beers. Furthermore, we note that the own-price elasticities of imported beers are also higher than for domestic beers, a pattern we already noted in the initial data exploration.

Hypothesis 3: Do promotions work better for foreign beers, since they are less well known? To study this hypothesis, we included interaction variables that tell us the difference between the effect of a promotion for a foreign brand from the effect of a promtion for a domestic brand. The estimated coefficients on these variables are negative and significant throughout, suggesting that contrary to our expectation, promotions work less well for foreign brands than for domestic brands.

Bonus hypothesis: Beer tastes better in summer. Although not reported in table 4, we also included month dummies in our regressions, and notice particularly postive and significant coefficients for the months of June and July, possibly revealing information about increasing demand due to holiday seasons and higher demand for a cool beer on a hot summer night.

# 5    Conclusion

In this report we use Scanner data from the 1990s collected in the Chicago area to study the demand dynamics for beer. Initial data exploration informs our hypothesis that demand in the beer market is best described by taking into account substitution patterns within and across two main segments; imported and domestic beer. We then lay out our modelling framework, which constists of step-wise increases in model complexity. We begin by fitting a simple multiplicative homogenous demand model. At this level, we find that demand for beer as a homogenous good is inelastic, but that this model performs relatively poorly in explaining demand variation.

We instrument for the retail price using the wholesale price, and further purge the wholesale price of any potential endogenous variation using a hedonic regression. We test the instrumented retail price resulting from this first stage in the homogenous demand model, and then proceed to use it for the further models too.

To account for finer demand dynamics in the beer market, we proceed to fit a multinomial logit and a nested logit demand model. The explanatory power of these models is significantly better than the homogenous demand model, and they allow us to credibly estimate the substitution patterns that we are interested in. Interestingly, both models return relatively high own-price elasticities, which are higher than we expected from our reading of the literature. The Nested Logit model further allows us to compare the elasticities of between and within group substitution. From our models we conclude that substitution between foreign and domestic beer is relatively low, whereas substitution within groups relatively higher. This lends support to our first hypothesis, namely that consumers do not tend to switch between group as much as within groups.

We include promotions and advertising as explanatory variables in the multinomial and nested logit models, to learn about their effect on beer demand. We find that general price promotions are the most effective, followed by coupon promotions and advertising. However, we cannot infer that promotions are also cost-effective, beacuse we do not have data on the cost of the promotions and advertising we observe.

Lastly, we also included an interaction variable between promotions and advertising with a group dummy for foreign vs domestic beers. This allows us to study whether the effect of promotions is higher for foreign than for domestic beers, assuming that the size of promotions indicated by our dummy is comparable. To our surprise, we find that the effect of promotions for foreign brands is lower than for domestic brands. We expected the opposite, given that foreign brands have less recognition value. However, it is possible that promotions for foreign brands are significantly smaller and therefore there effect is smaller, but our dataset does not allow us to investigate this.

# 6    References

Berry, S. T. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics 25(1994):242–262.*

Toro-González, D., McCluskey, J., and Mittelhammer R. "Beer Snobs Do Exist: Estimation of Beer Demand by Type" *Journal of Agricultural and Resource Economics 39-2(2014):1–14*

# 7 A/B Testing

Website: El Corte Ingles

## 7.1 Business Model:

El Corte Ingles is the 4th largest department in Europe and the only one remaining in Spain. They sell a wide plethora of products including music, electronics, cars and real estate. They operate a customer-centric business model and state customer satisfaction as their top priority. Online presence is an important part of their business, with 227 million visits registered to their website in 2014, an increase of 46.7% on the previous year. In 2016, they reported an increase of 40% in turnover from the website. Evidently, a significant proportion of their revenues are driven by online sales and hence an optimal website design is important for maximizing customer conversion rates.

## 7.2 Hypotheses:

- Increasing the number of items displayed on the primary banner on the website will lead to an increase in sales/conversion rate:
  - We regard three items on the primary banner to be too few and think that El Corte Inglès could generate more sales from displaying more products in this key section of the website.
  - Their current design requires extensive scrolling to view the full contents of the first page, we believe reducing this could be effective for retaining customers' attention.

- Removing price labels from the welcome page adverts will increase click through rate and lead to more conversions/sales:
  - We believe that some consumers will not click on an advert as the displayed price is too high for them initially.
  - By removing this price initially, you may convert some of these customers if they learn more information about the product after clicking through.

- Displaying available categories as an obvious banner at the top of the page instead of inside a drop-down menu will increase click through rate and the amount of time each customer spends on the website, and hence increase sales:
  - We believe that not having ready access to the available categories will turn away some customers.
  - Making the available categories clear and evident should capture some of these lost potential customers.

## 7.3 Testing Procedure

### 7.3.1 Hypothesis 1:

To test for the first hypothesis, we construct an A and a B scenario where A contains the original page format, B contains the same layout but with five products on the primary banner instead of three. These extra two products will be chosen by the same algorithm that produces the initial three. These two scenarios are then randomly assigned to each user of the website when they navigate to the page. The maximum amount of metrics should be captured, in order to understand the most about this change; but surely we would want to capture click-through rates (number of clicks of the ads relative to visits), click-through probability (number of unique visitors who click at least once relative to the number of unique visitors), and conversion rates (number of sales/click).

### 7.3.1.1 How to assign treatment and follow user interactions:

To capture as many metrics as possible, we want to follow the user's browsing throughout the session, and ideally multiple for multiple sessions. The ideal way to do this would be to follow users based on username. However, we expect many customers at El Corte Ingles do not have a customer account and if they do, might not usually be logged in. Therefore we presume it better to assign customers to A or B groups using cookies. This also has limitations, as Cookies can be deleted or the customer could change browser. We hope to get a large sample size for this experiment that should hopefully mitigate these effects, assuming the attrition from the sample is random and not correlated to group assignment (i.e. we would have to assume users that are assigned to group A with three products in the header are not more likely to delete cookies than users assigned to group B with five product in the header).

### 7.3.1.2 Sample Population and Duration:

Running the experiments for 1 month should be more than sufficient as the website receives around 19 million visitors/month (2014 numbers), and would allow us to observe within week 'seasonality'. To avoid a situation where users know that they are part of an experiment (by for example sitting next to each other and browsing the website together), we could segment the experiment geographically, making sure that the locations are very similar. Splitting by location depends on the ability of El Corte Ingles to identify user location - which we are not sure about. Within the two locations chosen, we would suggest using the whole user population, rather than focusing on subpopulations. The design choice under question is general enough to affect all users, so we would not want to miss a negative effect on one demographic by excluding them from the experiment.

The actual sample size required for the experiment can be calculated based on the expected size of the effect on our metrics. The smaller the effect we expect, the larger the sample required to estimate it reliably (we need enough pwoer). In this case, we expect the effect to be relatively small but still important.

### 7.3.1.3 Robustness:

Given that we expect a relatively small effect, we want to be sure our estimates are robust, and sugges bootstrapping the standard errors to produce more conservative estimates of the error. For this we resample the data many times with replacement and use this to compute sample statistics. If our sample size is not large enough to allow bootstraping standard errors, we would need to assume a distribution (Bernoulli in the case of click-through) which results in the pooled variance $SE_{pool} = \sqrt{(p(1-p)/N)}$; this tends to underestimate the standard error.

Using bootstrapped values for the standard error will give a lower Type 1 error; hence we will be more conservative with our reported significances.

### 7.3.2 Hypothesis 2:

To test for this hypothesis, we would run a similar experiment to the one described above where A contains the original page format and B contains the original page with the prices removed from the adverts.

We then randomly assign each of these two scenarios to each user who visits the page. We will evaluate the same metrics, with particular focus on click-through rate and conversion rate. As opposed to the change proposed under the first hypothesis, we expect the removal of price tag to have elicit much more heterogeneous responses. This means in the setup of the experiment we would place high importance of accurate identification of the customer demographic, so it might make sense to restrict the experiment only to logged-in customers, though then we would not learn whether the effect differs between regular and new customers. It might make sense to first test this for existing customers, learning about heterogeneity, and then decide whether to expand the experiment to new customers. It is possible that the optimal design consists of showing prices to some customers, and not to others, based on what we know about the customer. To optimise design in this case would therefore likely require a number of follow-up experiments.

The above considerations for assigning the treatments, choosing the populations, size/duration and robustness would all apply here.

### 7.3.3 Hypothesis 3:

Again, a similar situation as the two above hypotheses with A remaining as the original page and B presenting the home page with available categories as a banner at the top. We would focus mostly on click-through rate on categories as a whole here to measure the effectiveness of the treatment. For this experiment, we would be particularly interested in following the customer throughout the whole session, because we don't want the expansion of the category prominence on the front page to cannibalise other customer conversion channels.

### 7.3.4 Notes on multiple metric testing and testing of hypotheses:

For all of our hypotheses we would use the null that the control and treatment are identical:

$$H_0 : d = 0 \quad and \quad H_1 : d \neq 0 \quad where \quad d = p_{treat} - p_{cont}$$

where $p$ could represent click-through rate for example.

Then we work under the assumption that under

$$H_0 : d \sim N(0, SE)$$

[Central Limit Theorem] where we use the pooled standard error definition.

As we have a two-sided hypothesis test, we will reject if the estimated

$$d > Z * SE_{pool} \quad or \quad d < -Z * SE_{pool}$$

where $Z$ is set based on our required significance level alpha.

When we are testing multiple metrics, you increase the chances of obtaining a significant result, hence we must account for this by scaling our required significance level. We would use a Bonferroni correction

$$\alpha_{overall} = \alpha_{individual}/n$$

with an $\alpha_{overall} = 0.05$. So if we were testing 3 separate hypotheses for 3 different metrics, we would test each at the significance level 0.05/3.