A Random Coefficients Model of Seafood Demand: Implications for Consumer Preferences and Substitution Patterns

Kevin D. Ray, *Pacific States Marine Fisheries Commission*; Daniel K. Lew, *Alaska Fisheries Science Center*; Rosemary Kosaka, *Southwest Fisheries Science Center*

ABSTRACT

Discrete choice models of demand are growing in popularity for understanding markets for seafood, but have thus far been limited to applications using individual-level choice data. The random coefficients logit model is a discrete choice demand model designed for aggregate sales data and imparts a number of theoretical and empirical advantages. Instrumental variables account for price endogeneity, which can arise when there are unobserved product characteristics. Furthermore, correlated preferences can be accommodated in the random coefficients as well as through demographic interactions, which is especially important for seafood where product characteristics are primarily qualitative. We estimate this model for salmon fillets using four years of county-level seafood sales in California, and demonstrate the insights that can be drawn regarding consumer preferences and substitution patterns. Although the model is computationally burdensome, it offers considerable potential for further seafood demand analysis.

Key words: Consumer heterogeneity, cross-price elasticities, differentiation instruments, diversion ratios, price endogeneity.
JEL codes: D12, Q21, Q22.

INTRODUCTION

The primary goal of most seafood demand studies is to identify price elasticities or consumer willingness to pay for characteristics such as wild-caught, ecolabeled, or locally produced. The price elasticities are important for addressing concerns such as the impact that increased aquaculture production may have on wild-capture fisheries, or how temporarily shutting down a fishery might impact the market. The willingness to pay measures can help the industry in their decisions of how and what to produce. (For instance, do ecolabels make sustainable fishing practices profitable?) With efforts to develop and promote aquaculture in the United States and globally (NOAA Fisheries 2019; FAO 2020), and the recognition that media attention and industry/nongovernmental organization communications about both positive (e.g., health benefits and sustainability

Kevin D. Ray is under contract with the Pacific States Marine Fisheries Commission in support of the Southwest Fisheries Science Center, NOAA, National Marine Fisheries Service, 110 McAllister Way, Santa Cruz, CA 95060 USA (email: Kevin.Ray@noaa.gov). **Daniel K. Lew** is an economist, Alaska Fisheries Science Center, NOAA, National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle, WA 98115 USA (email: Dan.Lew@noaa.gov). **Rosemary Kosaka** is an economist, Southwest Fisheries Science Center, Santa Cruz Laboratory, NOAA, National Marine Fisheries Service, 110 McAllister Way, Santa Cruz, CA 95060 USA (email: Rosemary.Kosaka@noaa.gov).

The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce. We are grateful to Kristy Wallmo for feedback on an early draft, as well as the anonymous reviewers and associate editor, whose feedback has improved the paper.

Received January 5, 2022; Accepted September 26, 2022; Published online February 22, 2023.

Marine Resource Economics, volume 38, number 2, April 2023.
 https://doi.org/10.1086/723730

practices) and negative (e.g., ecological damage, mercury content) aspects of seafood production and consumption affect consumer perceptions (Amberg and Hall 2008) and preferences (Teisl, Roe, and Hicks 2002), there will be plentiful cause to estimate seafood demand in the coming years.

Demand studies for seafood, and salmon in particular, have evolved with improvements in econometric techniques and computing power. The earliest studies used single equation inverse-demand estimation to estimate the price elasticity of demand (Kabir and Ridler 1984; Bjørndal, Salvanes, and Andreassen 1992; DeVoretz and Salvanes 1993). The next step forward came with the development of demand system models, most notably the Almost Ideal Demand System (AIDS) introduced in Deaton and Muellbauer (1980). This model allows for multiple products to be included in the demand system, and with careful parameter constraints imposed, it conforms to the requirements of utility theory. Other demand system models such as Rotterdam and Leontief have been infrequently used in seafood demand, and in the case of Xie, Kinnucan, and Myrland (2009), hypothesis tests reject all such models except AIDS. Salmon demand systems estimated using AIDS typically employ time series data, with the earliest such study of Wessells and Wilen (1993) correcting for autocorrelation, while subsequent papers put significant focus on testing for unit roots and cointegration to produce reliable results (Asche 1996; Asche, Bjørndal, and Salvanes 1998; Xie and Myrland 2011; Singh, Dey, and Surathkal 2012). With few exceptions, these studies find that fresh salmon is price elastic and a normal good, while frozen salmon is often price inelastic and even an inferior good in one case (Asche, Bjørndal, and Salvanes 1998).

Unfortunately, the fashion in which the AIDS model handles products can be a major drawback. The model suffers from the curse of dimensionality; it grows in complexity with the square of the number of products. In order to have a tractable model, one must use aggregated definitions for the products. This aggregation is tied to market integration, so the generalized composite commodity theorem of Lewbel (1996) is used to analyze the appropriateness of product aggregation (for a thorough treatment, see Asche, Bremnes, and Wessells [1999]). However, the results can vary on the basis of how products are aggregated. For instance, the same data are used in Asche (1996) and Asche, Bjørndal, and Salvanes (1998), but the former aggregates salmon products into fresh, frozen, and smoked salmon, while the latter aggregates products into fresh Atlantic, frozen Atlantic, and frozen Pacific salmon. Looking at the frozen products, the former paper estimates the price elasticity for aggregate frozen salmon to be –0.28 with an expenditure elasticity of 0.16, while the latter paper finds that frozen Atlantic salmon has an own-price elasticity of –1.86 and an expenditure elasticity of 2.73 and that frozen Pacific salmon has price/expenditure elasticities of –0.51 and –0.27, respectively. Disaggregating the frozen salmon based on species/origin radically changed the results, with one inelastically demanded normal good becoming two goods: one an elastically demanded normal good and one an inelastically demanded inferior good. With increasingly detailed UPC-level product data becoming available to researchers, the potential issues from aggregating the data into a tractable number of products for AIDS become more concerning.

One solution to this issue is to consider the demand not as a function of products but as a function of a bundle of characteristics (Lancaster 1966). Building on the multinomial logit model (MNL) of McFadden (1973), there have been several salmon demand studies employing the random parameter logit (RPL) model, which is flexible enough to approximate any random utility model (McFadden and Train 2000). This model avoids the curse of dimensionality by considering the impact of product characteristics on utility, and it avoids the restrictive substitution patterns imposed by the independence of irrelevant alternatives (IIA) property of the MNL and nested logit models by allowing for random preference variation across consumers. In contrast to the AIDS studies that use international trade data and retailer scanner panels, the RPL model requires individual consumer choice data.

With respect to salmon demand, the RPL model has been used to gain insights into consumer preferences for labels such as organic and sustainably harvested.[1] Bronnmann and Asche (2017) surveyed shoppers at retail outlets in Germany for their choice of salmon products with varying price, production process, sustainability certification, and processing. They find substantial consumer heterogeneity and, on average, positive willingness to pay for sustainability certification. Respondent demographics are interacted with the alternative specific constant, which means that they can only influence the likelihood of a consumer choosing the "neither of these products" option. Because few of these demographic coefficients are statistically significant, the authors suggest that demographics do a poor job of capturing consumer preference heterogeneity. However, without interacting demographics with the product characteristics, they cannot identify whether a demographic variable influences preferences over the characteristics; for example, whether high-education consumers prefer certified sustainable salmon to uncertified salmon. The RPL is also used to explore ecolabel preferences in Ankamah-Yeboah et al. (2020), in which a consumer scanner panel is used to estimate consumer preferences for salmon characteristics including various processed forms, sizes, and organic and Marine Stewardship Council (MSC) labels. They find that the organic and MSC labels are not preferred on average, but the random parameters show that 26% of consumers prefer organic and 33% prefer MSC.

While the RPL model has been used numerous times for seafood with both choice experiments and consumer scanner panel data, the analogous random coefficients logit of Berry, Levinsohn, and Pakes (1995, abbreviated BLP), which is designed for aggregate sales data, has not yet been used for seafood demand modeling to our knowledge. The model has other advantages in addition to accommodating aggregate sales data. The major advantage is that it allows for endogeneity of prices, which can arise when there are product characteristics unobserved to the econometrician that are known to the producer/retailer and incorporated into the price, which is likely to be the case in any data that do not originate from a controlled choice experiment, such as what shade of pink/red the salmon is. Second, the method makes it straightforward to include market-level demographics in interactions with the product characteristics and the constant; this approach can provide richer results and allows for a structured correlation of preferences across characteristics. Another helpful feature is that, given the form of the model, it is not as susceptible to the issues of unit roots, which have dominated the discussion of AIDS models for salmon. The transformation of the problem into utility space, the use of a structural error, and the instruments for price all help to alleviate standard time series analysis concerns.

Although the BLP and RPL models essentially perform the same function, the BLP model, with its origin in the industrial organization literature, has often been used for questions regarding price-cost margins, models of oligopolistic competition, and merger simulations. This is in

1. The hedonic price approach has also been used for a similar purpose, estimating the market value of seafood characteristics. However, it should be noted that the hedonic price approach measures the equilibrium between consumer willingness to pay and producer willingness to accept.

contrast with the RPL literature, which frequently focuses on the estimated coefficients, willingness to pay, and consumer welfare. The seminal paper of Berry, Levinsohn, and Pakes (1995) analyzes the market for automobiles in the United States and demonstrates the model's ability to produce logical substitution patterns and insightful parameter estimates. One of the next major papers to use the model is Nevo (2001), which analyzes the ready-to-eat breakfast cereal market and demonstrates the effectiveness of demographics in explaining taste heterogeneity in aggregate data and the more reasonable cross-price elasticities that result from the flexible substitution pattern (it does not force independence of irrelevant alternatives), and finally concludes that a Nash-Bertrand pricing game is consistent with the observed price-cost margins.

The goal of our study is to demonstrate the applicability of the BLP model to seafood demand using a unique dataset of fresh and frozen seafood purchases from California. One feature of seafood that is unusual for BLP model estimation is the lack of any continuous covariates other than price; only categorical variables are present in the data. Nevertheless, the BLP model is still appropriate and significantly more flexible than a multilevel nested logit. Using the most recent advances in terms of the instrumental variables, our results show that product condition (i.e., fresh or frozen) is the most important characteristic in terms of shaping the substitution patterns. This result is logical given that these products are generally in different sections of the store. Furthermore, we find that a higher income is associated with a greater preference for salmon fillets compared with other salmon and seafood products.

The remainder of the paper proceeds as follows: First we describe the random coefficients logit methodology. Next we describe our dataset, followed by a description of the model specification and the configuration options used for the estimation. The results are presented and discussed, followed by our concluding remarks.

METHODOLOGY

We present a brief summary of the BLP model and solution algorithm, and recommend that those seeking a more thorough treatment see Nevo (2000) and Conlon and Gortmaker (2020). We observe sales in $t = 1, \ldots, T$ markets, defined as a county-quarter pair, each with $i = 1, \ldots, I$ artificial consumers. In each market we observe the average prices and aggregate quantities for J products, each with a set of known characteristics. Because we have aggregate sales data, we do not actually observe individuals or their choices, but we generate I individuals, each with a vector of random taste parameters and demographic characteristics based on distributions of these characteristics from the Census Bureau's American Community Survey data. For simplicity, we will omit the "artificial" when referring to consumers.

The indirect utility of consumer i from consuming product j in market t is defined as

$$u_{ijt} = \alpha_i \left(y_i - p_{jt} \right) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt}, \tag{1}$$

where y_i is the income of consumer i, p_{jt} is the price of product j in market t, x_{jt} is a K-dimensional row vector of observed characteristics of product j, ξ_{jt} is the mean market level utility of the unobserved characteristic(s), and ε_{ijt} is a mean-zero stochastic term with a type I extreme value distribution. The coefficient α_i is consumer i's marginal utility of income (or marginal disutility of price), and β_i is a K-dimensional column vector of individual-specific taste coefficients. This formulation assumes that there are no wealth effects from the decision to purchase salmon, consistent with other studies of grocery products (Nevo 2001; Villas-Boas 2007).

We can further break down the random coefficients α_i and β_i into their averages, that which can be explained by observed demographics, and that which is correlated with unobserved characteristics:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma \nu_i, \quad \nu_i \sim P_\nu^*(\nu), \quad D_i \sim \widehat{P_D^*}(D), \tag{2}$$

where D_i is a $d \times 1$ vector of demographic variables, ν_i captures the additional unobserved preferences, and P_ν^* is a parametric distribution of random draws (in our case, multivariate normal) and $\widehat{P_D^*}$ is the distribution of demographic variables. The matrix Π contains the coefficients that measure how tastes vary with demographics, and the matrix Σ is the set of random coefficients. Specifically, Σ is the Cholesky root of the preference covariance matrix. The set of parameters to be estimated are labeled θ, with θ_1 containing the linear parameters (α and β) and θ_2 containing the nonlinear parameters (Π and Σ).

Typically Σ is assumed to be diagonal, limiting the cross-characteristic preference correlations. Although this is a restrictive form, it is common in the random parameters logit literature. This is critiqued for the RPL model in Mariel and Meyerhoff (2018), with a recommendation to allow for correlation and test the hypothesis that the restricted model is valid. In the random coefficient model, the demographic interactions of the Π matrix actually allow for a form of structured correlation of preferences. For instance, based on the results of Nevo (2001), high-income consumers are more likely to purchase a cereal that gets soggy in milk and less likely to purchase a sugary cereal. The inclusion of both these parameters in the Π matrix means that, although it still assumes zero correlation for the random coefficients in Σ, there is some correlation in preferences that will appear in the estimated substitution patterns.

Combining equations 1 and 2 completely describes the model, and it becomes apparent that it can be simplified by collecting the elements that vary by individual, $\mu_{ijt} = [-p_{jt}, x_{jt}](\Pi D_i + \Sigma \nu_i)$, and those that are fixed across individuals. These fixed elements are the market-specific product utility averages, $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}$, with ξ_{jt} being the structural error term that represents the mean market-level utility of the unobserved characteristics. Finally, the demand system is completed by the addition of an outside good, for which the standard practice is to normalize the utility to zero.

Consumers are assumed to purchase the good that gives the highest utility (inclusive of the outside good). This implies that the products are all substitutes, which seems to be a reasonable assumption for salmon fillets. Because we do not observe actual consumers and purchases, but rather a vector of random taste preferences, demographics, and product-specific errors, we construct the set of consumers choosing product j in market t as

$$A_{\{jt\}\left(x_{\{\cdot t\}}, p_{\{\cdot t\}}, \delta_{\{\cdot t\}}; \theta_2\right)} = \left\{\left(D_i, \nu_i, \varepsilon_{\{i0t\}}, \ldots, \varepsilon_{\{iJt\}}\right) | u_{\{ijt\}} \geq u_{\{ilt\}}, \quad \forall\, l = 0, 1, \ldots, J\right\}. \tag{3}$$

Integrating over this set will recover the market share, which suggests that the estimation procedure should essentially select parameters that lead to predicted market shares that are close to the observed market shares. Because of the endogeneity of price, this is accomplished by an instrumental variables generalized method of moments (GMM) approach with a matrix of instruments Z'_{jt} and weighting matrix W, and seeks to solve

$$\min_\theta \left(\frac{1}{N}\sum_{j,t} Z'_{jt}\xi_{jt}(\theta)\right)' W \left(\frac{1}{N}\sum_{j,t} Z'_{jt}\xi_{jt}(\theta)\right). \tag{4}$$

The solution algorithm begins with an initial value for θ_2 and then estimates the mean market-level utilities δ_{jt} with a contraction mapping. Next the objective function value is computed. The nonlinear optimizer selects the next values of θ_2 for estimation and the process repeats until convergence.

INSTRUMENTS

The instruments are a critical element of the model, and although price is the only variable we assume to be endogenous, an instrument is needed for each element of θ. Although early papers focused on finding instruments that are excludable and relevant for price, such as the Hausman type instruments in Nevo (2001) and the input-cost instruments in Villas-Boas (2007), the instruments are actually identifying all of the parameters and therefore their relevance for nonprice preference variation requires consideration. Gandhi and Houde (2019) demonstrate that good instruments for the BLP model can be created from the degree of differentiation between products in a market. The variant we employ is the *local* differentiation instruments, which are defined as the number of other products with a characteristic within one standard deviation of a given product. In the case of categorical variables, it is the number of products with the same characteristic. To better account for correlation in preferences, these instruments can be extended with interactions, in which one product characteristic value is multiplied by the number of products with another characteristic. Furthermore, in the case of endogenous prices, they recommend including a predicted price from a regression on observed characteristics, which can also be interacted. Evidence suggests that these are the best set of instruments that can be generated to solve the initial problem, but Conlon and Gortmaker (2020) suggest that the problem can be more precisely identified by using the approximation to the optimal instruments described in Reynaert and Verboven (2014) based on the results of the model with differentiation instruments.

ELASTICITIES AND DIVERSION RATIOS

The price elasticity of product j with product l in market t is given by

$$\eta_{jlt} = \frac{\partial s_{jt} p_{lt}}{\partial p_{lt} s_{lt}} = \begin{cases} {}^{-p_{jt}}\!/_{s_{jt}} \displaystyle\int \alpha_i s_{ijt}\left(1 - s_{ijt}\right) d\hat{P}_D^{*}(D) dP_{\nu}^{*}(\nu) & \text{if } j = l, \\ {}^{p_{lt}}\!/_{s_{jt}} \displaystyle\int \alpha_i s_{ijt} s_{ilt} d\hat{P}_D^{*}(D) dP_{\nu}^{*}(\nu) & \text{otherwise,} \end{cases} \tag{5}$$

where $s_{ijt} = \exp(\delta_{jt} + \mu_{ijt}) / [1 + \Sigma_{l=1}^{L} \exp(\delta_{lt} + \mu_{ilt})]$ is the probability of individual i purchasing product j. As with the RPL, the own-price elasticity is no longer determined by the functional form of the price variable, and the cross-price elasticity depends on more than just the market shares as a result of relaxing the IIA assumption. Note that there is a different elasticity for each market, so we will follow Nevo (2001) and present the median estimated own- and cross-price elasticity across all markets.

The basic BLP model with random coefficients on a single categorical variable will approximate a nested logit with different nesting parameters for each category. With respect to the elasticities, this model will have only two values per column, one for products that are in the same nest and one for products in any other nest. The addition of a second categorical variable results in a model that is similar to a multilevel nested logit; however, it is more flexible because it does not impose an order on the nests. In this model there would be only four different elasticity values per column, depending on whether the products share both nests, one of the two nests, or no nests. It is the addition of correlated preferences, either by allowing the random coefficients to be correlated or by including demographic interactions, that allows for much greater flexibility in the estimated elasticity values within a column. The correlated preferences fill the important role of improving the model's guess of the "second choice" product, which is what leads to the increased flexibility.

However, cross-price elasticities may not give us the most accurate picture of substitution patterns in the market. As is pointed out in Conlon and Mortimer (2021), the presence of the quantity in the elasticity calculation means that a product with a larger cross-price elasticity but a low market share could see fewer customers switch into it than a product with a smaller cross-price elasticity and greater market share. They suggest that the diversion ratio is an intuitive choice to identify these substitution patterns. This statistic addresses the question of how much consumers increase purchases of another product in the case that a price increase in one product spurs a reduction in purchases of that product. Mathematically the diversion ratio is given by $D_{jk} = \frac{\partial q_k}{\partial p_j} / \left| \frac{\partial q_j}{\partial p_j} \right|$. Given the question the statistic is designed to answer, it should not be surprising that an identity exists linking it to the ratio of the own-price elasticity (how much consumption of a product changes with its price change) and the cross-price elasticity (how much consumption of other products changes). Specifically, the diversion ratio $D_{jk} = -\frac{\epsilon_{jk}}{\epsilon_{jj}} \times \frac{s_k}{s_j}$, where ϵ is the elasticity and s is the market share. As with the elasticities, different diversion ratios are estimated for each market and therefore the median will be presented. Unfortunately this means that the sum of diversion ratios for a given row will not necessarily equal unity, but we believe the gains to including all markets are sufficient to justify this choice.

DATA

The data include county-level weekly grocery store sales of seafood products (UPC level) from California from January 2013 to December 2016 from FreshFacts, a division of Nielsen focused on fresh foods such as produce and meat. For salmon products, the data include quantities and revenues, as well as specifying the brand, species, production method, origin, product form, and a 30-character product description. For other seafood products, the data contain only quantities and revenues, but this is sufficient to represent consumption of the "outside good." It is a rare opportunity with these data that the market share of the outside good is observed rather than imputed based on things like per capita consumption or recommended servings.[2]

The observed product characteristics were used in concert with the product descriptions to improve the data. For a full description of the data cleaning and extrapolation process, see the online appendix to Ray, Lew, and Kosaka (2022). Contradictions were corrected, for instance Alaskan-origin Atlantic salmon was reclassified as generic origin[3] Atlantic salmon. Similarly, missing product characteristics were extrapolated from industry knowledge and data relationships. For instance, Alaskan-origin salmon is all wild-caught. The product descriptions contained some information regarding fresh, frozen, and previously frozen that were used to generate an additional variable, although this information was unavailable for 60% of the market.

2. For example, recent studies have found that Americans consume 2.4 pounds of salmon per year, and the FDA recommends adults consume 2 pounds of seafood per month. These paint a very different picture of the market, as we observe per capita annual purchases of seafood at 1 pound, and salmon at 0.4 pound.

3. Generic origin indicates that the origin is unknown to the researchers. We chose to use the term "generic" to avoid ambiguity with the "unknown" product condition.

Table 1. Summary of Included and Excluded Products

Product Type	Frequency (%)	Market Share (%)	Number of Products
Included salmon	33.0	34.0	360
Excluded salmon	40.1	18.1	783
Other seafood	26.9	47.9	N/A

Note: Other seafood product identifiers were masked by Nielsen, preventing count statistics from being computed.

Unfortunately, several common characteristics included in other demand studies are not available in these data. Package size is not included in the data, which is relevant for the prepackaged products but not the variable weight products, and will therefore be a part of the structural error (unobserved characteristics, possibly correlated with price). Ecolabeling information is also missing, although in this market context it is just a minor concern. The first salmon farms were certified by the Aquaculture Stewardship Council in 2014, partway through our study period, and if any of these products are included in the data then the ecolabel will be part of the structural error. However, the Alaskan salmon fisheries have gone through a complicated on-again, off-again relationship with the Marine Stewardship Council (Foley and Hébert 2013). Because all Alaskan salmon fisheries were certified, decertified, and recertified together, the place of origin label should contain the sustainability premium. The fact that the recertification took place in November of the first year of our sample, in addition to the fact that the MSC label was placed on only 5% of salmon sold in the United States, suggests that perceived differences in the two ecolabels are unlikely to significantly bias our results (Knapp, Roheim, and Anderson 2007).

The BLP model treats every product that is sold in a market as part of the consideration set, but that also means that any product with zero market share is assumed to have been out of the

Table 2. Summary Statistics

Product Attribute	Frequency (%)	Market Share (%)	Number of Products
Species			
Atlantic	38.1	67.0	104
Chinook	7.7	5.1	57
Chum	5.9	2.7	37
Coho	7.6	4.1	60
Sockeye	40.7	21.1	102
Origin			
Alaskan	45.0	23.3	137
Copper River	2.5	2.1	22
Generic	50.1	70.2	197
Norwegian	2.4	4.3	4
Production method			
Farmed	40.0	68.2	114
Wild	60.0	31.8	246
Condition			
Fresh	25.8	31.2	70
Frozen	9.4	4.9	50
Previously	10.7	3.2	18
Unknown	54.1	60.7	222

consideration set. This can lead to bias in high-frequency data (Dube, Hortaçsu, and Joo 2020), so we aggregate the data to the quarterly level to alleviate this concern and to reduce the computational burden, giving us a total of 640 markets.[4] At the quarterly level, the model still accounts for the seasonal availability of fresh Pacific salmon.

In order to improve stability of the model and reduce the computational burden, we further restrict the set of salmon products in the model. The excluded salmon products are not removed from the data, but rather become a part of the outside good. We include only fillet products, as these make up 82% of the market by value. The model was highly unstable with the inclusion of some products with rare characteristics, particularly pink salmon (0.6%) and salmon with origins in Chile (0.3%) and Scotland (<0.01%), therefore these products are also excluded. The final set of products to be excluded were those that appeared in only one county for a given quarter, since the Hausman price instrument could not be computed for these products. This results in a dataset of 360 included salmon products with a total market share of 34.0%, a market share of 18.1% for excluded salmon products, and 47.9% for all other seafood (see table 1). A summary of the relative frequency of product characteristics is presented in table 2.

SPECIFICATION AND CONFIGURATION

COEFFICIENTS AND FIXED EFFECTS

The model specification consists of the product formulation including the linear and nonlinear variables, and the agent formulation including the demographic variables. We allow the species, product condition (fresh/frozen), country of origin, and production method to enter the linear utility portion of the model.[5] Time fixed effects are also absorbed from the linear utility to control for any statewide changes over time.[6] For the nonlinear portion with random coefficients, we include only the species and product condition.[7] A specification with correlated random preferences (off-diagonal elements of Σ) with and without the demographic interactions was also tested, but the resulting coefficients were not individually statistically significant and produced minimal changes to the results, so we ultimately employed the simpler specification. For the agent formulation, we include the natural log of income and allow it to interact with all of the nonlinear variables, leading to a structured form of correlated preferences. In order to ensure that all consumers have a theoretically consistent negative coefficient on price, we impose a lognormal distribution on preferences for price (Carson and Czajkowski 2019).

We define the baseline product as farm-raised Atlantic salmon of generic origin and unknown product condition. This is primarily to leverage the model's handling of preference heterogeneity.

4. Comparing the count of products reveals that on average 27% of products sold in a given county-month had zero market share for some week within that month. The results comparing monthly with quarterly are similar, and suggest that the bias for high-frequency data described in Dube, Hortaçsu, and Joo (2020) would be an issue.

5. Interactions of product characteristics in this nonlinear portion of the model were tested, however the nonlinear search algorithm did not converge with these interactions included. This is likely specific to our problem, and such interactions are worth testing.

6. These single-dimensional fixed effects are absorbed by de-meaning (for each variable the quarterly average values are deducted) as recommended by Conlon and Gortmaker (2020) to reduce computational burden relative to including the full set of dummy variables.

7. The model becomes unstable with random preferences over production method and country of origin variables because of high levels of collinearity between species and these variables (see online appendix table A.1). This correlation also makes the coefficients more challenging to interpret because no ceteris paribus comparison can be made, for instance, between farmed Atlantic salmon and wild sockeye salmon because both species and production method differ across all observations.

Because the model computes the preference heterogeneity for all omitted characteristics together in the constant, one can learn very little about the preference heterogeneity for the individual product characteristics included in the baseline. Therefore we define the baseline product to include the nebulous product characteristics of generic origin and unknown product condition, and the most common species, Atlantic salmon.

INSTRUMENTS

In accordance with the procedure recommended in Conlon and Gortmaker (2020), we begin with the differentiation instruments of Gandhi and Houde (2019). To instrument for price, we adapt their procedure of estimating an exogenous predicted price $\hat{p}$ formed by taking the predicted values from a regression of product prices on characteristics and cost-shifters. Using the observed characteristics and the price of diesel and lagged price of fishmeal interacted with product dummies (similar to Villas-Boas 2007) does not adequately capture price variation, producing an adjusted R^2 of 0.45. Therefore, we supplement this model by computing a Hausman-type instrument defined as the average price of the same product in other counties for the same quarter and adding this to the regression, as well as interacting the diesel and lagged fishmeal prices with product dummies. This augmented model achieves an adjusted R^2 of 0.90. Given the improved fit, we proceed with this formulation to compute $\hat{p}$. The goal is to satisfy the conditional moment restriction $E[\xi_{jt}|\boldsymbol{x}_t, \widehat{\boldsymbol{p}_t}] = 0$, and although we have added terms to the regression approach recommended in Gandhi and Houde (2019), the Hausman-type instruments will still be independent of the market-specific valuation of the unobserved characteristics as required.[8]

In addition to this price instrument, we compute the local differentiation instruments for the price, species, and condition variables. For price, this is the count of products within one standard deviation of the observed product price. For the categorical variables, the resulting instruments are the count of products in a market with the same characteristic as a measure of the competition in that market niche. To facilitate the estimation of the structured correlation coefficients on the demographic interactions, the local instruments also include interactions. For the categorical variables, these are the sum of products with the same pair of characteristics (for instance, the count of fresh Atlantic salmon products). For the price variable, this takes two different forms: the count of products within a standard deviation of price with another characteristic, and the sum of prices of products with the same characteristic. This resulted in 80 differentiation instruments, of which 38 are perfectly collinear and must be removed, and 9 are highly collinear as measured by the variance inflation factor and are also removed, for a final total of 33 differentiation instruments plus the predicted price. This is more than sufficient to identify the 18 nonlinear coefficients.

We perform the independence of irrelevant alternatives hypothesis test described in Gandhi and Houde (2019) in order to establish the strength of the instruments and to verify that the ordinary logit is not sufficient to the problem. In essence, the test calls for estimating the ordinary logit model with the instruments included as independent variables. If the IIA assumption does not hold and the instruments are strong, then the hypothesis test that the instrumental variables' coefficients are jointly zero will be rejected. The computed F-statistic (33, 33984 df) is 73 with a p-value of 0.000, indicating that the instruments are strong and the ordinary logit is not appropriate.

---

8. We have tested the model with the simple Hausman-type instrument for price and the accompanying differentiation instruments. The results are not appreciably different from the results with the more complex predicted price instrument.

After utilizing the differentiation instruments described above to solve the random coefficients logit problem, the next step is to compute the approximation to the optimal instruments using these initial results. Conlon and Gortmaker (2020) and Reynaerts and Verboven (2014) show that the simple approximation to the optimal instruments based on the Jacobian computed at $E(\xi) = 0$ performs similarly to more computationally complex Jacobian computations based on averages computed over the normal or empirical distribution. We proceed to use this simple approximation to the optimal instruments, which is just-identified when no supply side is estimated (as is the case here).

QUASI-RANDOM DRAWS

As was mentioned previously, the model relies on numerical integration over a number of artificial consumers with randomly assigned preferences and demographics. Although early papers use pseudo-random draws for the random preferences, these are inferior to quasi-random number generators for both random parameters logit and BLP (Czajkowski and Budziński 2019; Conlon and Gortmaker 2020). The Sobol draws (a variant of $[t, m, s]$-nets) recommended in Czajkowski and Budziński (2019) do a particularly good job at providing even coverage of the domain in multiple dimensions with fewer draws than other methods. We chose to use 10,000 Sobol draws per market for our numerical integration.

For the income variable, we used the binned income distributions reported by county in the one-year American Community Survey for each year in the study period. These data are only published for counties with more than 65,000 population, so only 40 counties are retained in the analysis dataset. We feel that the benefits gained by using the most current income data outweigh the loss of data from the 12 smallest counties in the sample.[9] The binned income distributions were used to generate a series of 10,000 bin assignments for each market, with each of these bin assignments used to generate a numerical income through a uniform random draw over the range of each bin (with an arbitrary cap for the top-coded bin). The random draws from the two highest bins (150k–200k and greater than 200k) are then re-scaled to ensure that the mean for each county-year is the same as the American Community Survey observed county average.[10] The importance of preserving the sample mean is demonstrated in Von Hippel, Hunter, and Drown (2017), which uses the exact Gini coefficient calculated by the Census Bureau from unbinned data and estimated Gini coefficients from binned data to compare methods. They report that the estimates from all methods "improve dramatically" when the simulated distribution is constrained to match the published county means. With the goal of the quasi-random draws being to reproduce the population distribution, we believe this is an important step. We then take the log of this computed income value and de-mean the variable in order to preserve the zero expected value for the idiosyncratic error and the structural error. These income draws are combined with the Sobol draws to complete the agent data.

9. The remaining six California counties did not have any retailers that share data with Nielsen.

10. We initially attempted to re-scale only the top-coded bin to replicate the observed means, but this resulted in negative income values for some markets. A related finding, that to preserve the grand mean for county income would require a mean income for the top bin that is less than its lower bound, is reported to be the case for 4% of US counties (Von Hippel, Hunter, and Drown 2017, 644).

PROGRAMMATIC CONFIGURATION

The model is estimated in Python 3 with PyBLP (Conlon and Gortmaker 2020). The full model consists of two steps: in the first step, the differentiation instruments are used to generate an estimate of the parameters that are used to compute the approximation to the optimal instruments, and in the second step the problem is solved with the approximation to the optimal instruments. We chose to use the same settings for both steps. The problems are solved with two-stage GMM to allow for improvements in the estimated results due to the updated weighting matrix. The nonlinear optimization algorithm is BFGS with the convergence criterion defined as a projected gradient norm less than 1e-4. To evaluate whether the algorithm converged to a local optimum rather than a global one, a different optimization algorithm (the conjugate gradient method) and different starting points were used for the second stage with the approximation to the optimal instruments. The conjugate gradient algorithm arrived at the same optimum, but some of the different starting points arrived at other minima with larger objective values. This indicates the presence of local minima, but the results we present are consistent with the global minimum. The iteration routine chosen for solving the fixed-point computation of δ_{jt} is the Levenberg-Marquardt algorithm.

RESULTS

The results of the model computed with the approximation to the optimal instruments are presented in table 3,[11] with the first-stage results computed with the differentiation instruments included in the online appendix tables A.3–A.5. The mean utility levels are largely as expected, although some of the magnitudes are surprising. The results seem to indicate that the product condition is many times more important to consumers than other characteristics, with products in the data from which the product condition could be determined from the product description being much more valued. The results show large positive and statistically significant estimates on fresh and frozen, as well as a large but insignificant estimate for previously frozen. The estimates for the mean product condition preferences are ordered as expected, with fresh preferred to previously frozen and frozen. Wald tests indicate that the average preference for fresh salmon products is statistically different from frozen ($p = 0.001$), but previously frozen is not statistically different from the other two conditions ($p = 0.869$ for fresh, $p = 0.836$ for frozen). However, caution should be used in applying these results, as the conditions of frozen and previously frozen are only observed for sockeye products (online appendix table A.1).

On their face, the results for species are also somewhat surprising, with the high-value Chinook and sockeye estimated to have negative and insignificant mean utilities. However, the ceteris paribus comparison is misleading because all sockeye and most Chinook in the data are wild-caught in Alaska (see online appendix table A.1). To compare the average preferences for farmed Atlantic salmon and wild-caught sockeye and Chinook, we need to add the wild-caught coefficient of +1.66 and the Alaskan coefficient of –0.10 to the species coefficient. With estimated species coefficients of –0.97 for sockeye and –0.12 for Chinook, the sums are +0.59 and +1.44, respectively, indicating that these products are preferred on average to the baseline product of

11. We also computed a version of the model that did not de-mean the data and instead used dummy variables for the time fixed effects. This allows for the recovery of the constant in the mean utility, which indicates the relative preference between the baseline product and excluded products. These results are presented in online appendix table A.2.

Table 3. Results Computed with Approximation to Optimal Instruments

	Mean Utility (β)	Preference Heterogeneity (σ)	ln(Income) (π)
Constant		0.03	25.44*
		(1.55)	(2.61)
Negative price ($-p$)	−1.47*	0.23	−0.09
	(0.26)	(0.45)	(0.06)
Condition			
Fresh	26.01*	0.74	−22.77*
	(2.35)	(2.54)	(1.98)
Frozen	17.18*	0.81	−28.44*
	(3.05)	(0.99)	(2.78)
Previously frozen	22.14	2.51	−22.21*
	(21.58)	(12.41)	(3.52)
Species			
Chinook	−0.12	0.30	−1.50*
	(4.45)	(16.29)	(0.40)
Chum	−4.76*	2.23*	−0.11
	(1.02)	(0.58)	(0.14)
Coho	−5.47*	2.66*	−0.14
	(0.90)	(0.51)	(0.14)
Sockeye	−0.97	1.15	−0.51*
	(1.00)	(1.58)	(0.12)
Origin			
Alaskan	−0.10		
	(0.25)		
Copper River	0.42		
	(0.31)		
Norwegian	1.40*		
	(0.38)		
Production method			
Wild-caught	1.66*		
	(0.19)		

Note: The random preferences for price are lognormally distributed. This is implemented by including the term $(-p)e^{(\beta_p+\sigma_p\nu_p+\pi_p y)}$, where β_p is the lognormal distribution mean, $\sigma_p\nu_p$ is the product of the random coefficient and the individual heterogeneous preference draw, and $\pi_p y$ is the product of the interaction coefficient and the income draws. There is no constant in the mean utility because time fixed effects are absorbed through a de-meaning procedure. For more information see footnotes 6 and 11, or see online appendix table A.3. * indicates significance at the 95% level.

farmed Atlantic salmon. However, the Wald test for significance of these combined coefficients indicates that these are still insignificant effects ($p = 0.570$ for sockeye and $p = 0.746$ for Chinook). The negative and significant coefficients on chum and coho are consistent with hedonic price studies that compare the value of different salmon species (Asche et al. 2005; Asche et al. 2015), and remain significantly negative with the addition of the wild-caught coefficient. It is interesting to note that the Alaskan and Copper River origins are not statistically significant compared with the baseline product but are significantly different from one another ($p = 0.025$). The insignificance of origin is likely connected to the collinearity with species and production method described above. With Copper River origin making up such a small fraction of the data, it may just not be able to precisely estimate the value. However, imported Norwegian Atlantic salmon is clearly preferred to other sources of Atlantic salmon.

The mean utility level results can be used to compute an average willingness to pay for characteristics by taking the ratio of the characteristic marginal utility and the marginal disutility of price. Because of the lognormal distribution, the mean disutility of price is given by $e^{(\beta+0.5\sigma^2)} = e^{(-1.47+0.5\times 0.23^2)} = 0.236$. So with the aforementioned combined values for Alaskan sockeye (0.59) and Alaskan Chinook (1.44), this implies that the average willingness to pay for these products is higher than Atlantic salmon by \$2.50 for sockeye and \$6.10 for Chinook. This is similar to the observed average price difference for these species. The negative and significant coefficients for chum and coho remain negative and significant even accounting for the preference for wild-caught. This same procedure could be performed for the wild-caught characteristic alone, producing a figure that is much higher than what is observed in hedonic price studies ($1.66/0.236 = 7.03$). However, as mentioned above, this ceteris paribus comparison is problematic given the sparseness of characteristic combinations. This example highlights not only the procedure for computing willingness to pay, but also the importance of using it properly in the context of categorical variables.

The results for the nonlinear coefficients indicate that the unexplained heterogeneity plays a small role, with only chum and coho having statistically significant unexplained preference heterogeneity.[12] Meanwhile, most of the demographic interactions are statistically significant. This is not uncommon in BLP studies (e.g., Nevo 2001); however, Bronnmann and Asche (2017) note that the insignificance of demographic interactions in the RPL model supports Grebitus, Jensen, and Roosen (2013), which argues that demographic variables do not capture consumer heterogeneity well. This difference may be related to the use of individual consumer choice data as opposed to aggregate sales data, but it is important to note that Bronnmann and Asche only allow for demographic variables to interact with the constant. The type of data may matter, because when observing the decisions of an individual, unobserved individual heterogeneity is likely to be more impactful than sociodemographics. However, when measuring aggregate decisions, the aggregate demographics are more informative. For example, if many Japanese Americans have a preference for and purchase more salmon relative to other races, then the fact that many Japanese Americans shop in a given market will imply that more salmon will be sold. However, with individual data, observing several Japanese Americans who dislike salmon and never choose it may be sufficient to render the sociodemographic variable statistically insignificant.

The only product characteristics with statistically significant unexplained preference variation are the lower-valued species of chum and coho. Combined with the near-zero coefficients for these species in the demographic interactions, this indicates that the preferences for chum and coho salmon products vary across the population in ways that are not explained by income, and with the preferences for price allowed to vary in the model this indicates that these consumers do not simply prefer these characteristics due to a lower price.[13]

With numerous statistically significant parameters and relatively large magnitudes, the results for the log-income interactions indicate that this socioeconomic variable is important for explaining preference heterogeneity. For the coefficients with statistically insignificant unexplained heterogeneity and significant income interactions, this indicates that after accounting for income-driven preference heterogeneity, there is no significant unobserved Gaussian preference heterogeneity. The large and significant estimate for the interaction with the constant suggests that

12. Estimating the model without unexplained heterogeneity produces similar point estimates for the model coefficients, median price elasticities, and diversion ratios. However, a Wald test of the joint hypothesis that all of the coefficients for the unexplained heterogeneity are zero can be rejected ($p < 0.0001$).

13. Results were similar for models estimated with fixed preferences for price. The main change was generally larger standard errors with price preferences held fixed, which caused changes in significance of several coefficients.

higher-income individuals are more likely to choose salmon fillets than other seafood products (including other forms of salmon). Although the estimated price and income interaction is not statistically significant, since it is interacted with the negative of price, the negative sign is consistent with a decreasing marginal utility of income. The large and significant negatively signed estimated coefficients for all of the product condition attributes are surprising, but with the omitted category being "unknown" these estimates can provide limited economic interpretation. Inference can be made from the relative magnitudes, which reveal that higher-income households prefer fresh and previously frozen salmon to frozen, while lower-income households prefer frozen salmon. It is surprising that the two salmon species generally regarded as higher valued, sockeye and Chinook, have negative and significant income interactions. This suggests that higher-income households in California generally prefer Atlantic salmon.

The median price elasticities implied by the results for a selection of Atlantic salmon products are presented in table 4, which illustrates the change in quantity of the row product with a change in price of the column product. The particular products were selected from the products with the highest average market share among products that appeared in the most markets to minimize missing values where two products never coexist. The median own-price elasticities are generally similar to the results of prior AIDS analyses, with most products ranging from –1 to –2. However, the higher-price species of Chinook and sockeye have more elastic demand with estimates of –2.5 to –3.4. Compared with the aggregate product categories used in AIDS studies, the disaggregated product-level data are likely to produce more elastic estimates of demand because there are more substitutes with similar characteristics. To expand on this median elasticity analysis, a graphical representation of the distribution of own- and cross-price elasticities across all markets is presented in figures 1 and 2, with the median elasticity indicated by a +.

The cross-price elasticities demonstrate the primacy of the product condition effect as evidenced by the associated large coefficient estimates. Given that frozen salmon and fresh salmon are separated by some distance in most retail outlets, this is quite plausible. It is intriguing that the unknown product condition behaves like its own category entirely, but this could be easily explained if, for example, products described as "fresh" in the data are sold at the meat counter while those without the descriptors are sold in tray packs in the refrigerated meat section. What

Table 4. Median Price Elasticities for Selected Products

Species: Origin: Condition:	Atl Gen Fro	Atl Gen Fre	Atl Gen Unk	Chu Gen Fre	Coh Gen Fre	Chi AK Fre	Soc AK Fre	Soc AK Unk	Soc AK Fro
Atl Gen Fro	–1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20
Atl Gen Fre	0.00	–1.16	0.02	0.01	0.00	0.03	0.01	0.01	0.00
Atl Gen Unk	0.00	0.04	–1.47	0.00	0.00	0.00	0.00	0.08	0.00
Chu Gen Fre	0.00	0.20	0.01	–1.34	0.00	0.02	0.01	0.00	0.00
Coh Gen Fre	0.00	0.15	0.01	0.01	–2.06	0.01	0.01	0.00	0.00
Chi AK Fre	0.00	0.21	0.01	0.01	0.00	–3.37	0.01	0.00	0.00
Soc AK Fre	0.00	0.22	0.01	0.01	0.00	0.02	–2.67	0.01	0.00
Soc AK Unk	0.00	0.04	0.26	0.00	0.00	0.00	0.00	–2.54	0.00
Soc AK Fro	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	–2.74

Note: These are the median price elasticities across markets for the change in quantity of the row product with a change in price of the column product. Generic origin (Gen) indicates that Nielsen did not have reliable origin data and that the product description did not include that information. Unknown condition (Unk) indicates that the product description did not specify fresh, frozen, or previously frozen.

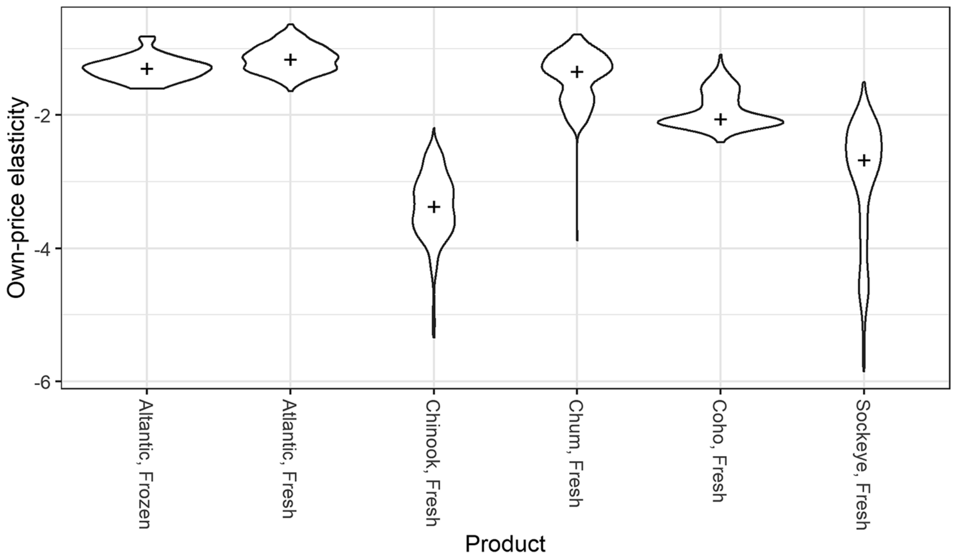

Figure 1. Distribution of Own-Price Elasticity Estimates for Selected Products

stands out from table 4 is the large cross-price elasticities in the fresh Atlantic salmon column, which is largely a result of the importance of the fresh condition in the model. The distribution of the cross-price elasticities in figure 2 demonstrate that the model's assumption of pure substitutability restricts the cross-price elasticities to the nonnegative domain and leads to asymmetric distributions.

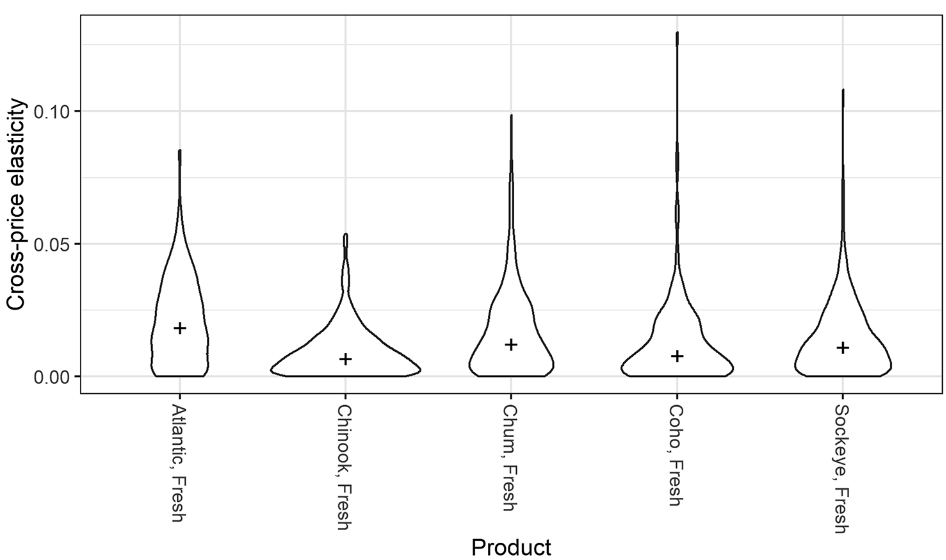

Figure 2. Distribution of Cross-Price Elasticity Estimates of Selected Products to One Atlantic Salmon Product

Table 5. Median Diversion Ratios of Selected Products

Species:	Atl	Atl	Atl	Chu	Coh	Chi	Soc	Soc	Soc
Origin:	Gen	Gen	Gen	Gen	Gen	AK	AK	AK	AK
Condition:	Fro	Fre	Unk	Fre	Fre	Fre	Fre	Unk	Fro
Atl Gen Fro	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08
Atl Gen Fre	0.00	0.74	0.01	0.01	0.00	0.01	0.00	0.00	0.00
Atl Gen Unk	0.00	0.03	0.05	0.00	0.00	0.00	0.00	0.03	0.00
Chu Gen Fre	0.00	0.15	0.01	0.56	0.00	0.01	0.00	0.00	0.00
Coh Gen Fre	0.00	0.11	0.00	0.01	0.58	0.00	0.00	0.00	0.00
Chi AK Fre	0.00	0.16	0.00	0.01	0.00	0.67	0.00	0.00	0.00
Soc AK Fre	0.00	0.16	0.01	0.01	0.00	0.01	0.59	0.00	0.00
Soc AK Unk	0.00	0.04	0.17	0.00	0.00	0.00	0.00	0.05	0.00
Soc AK Fro	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38

Note: The off-diagonal elements are the median diversion ratios across markets for the change in quantity of the row product with a change in price of the column product. The diagonal elements are the median diversion across markets to the outside good with a change in price of the column product. Generic origin (Gen) indicates that Nielsen did not have reliable origin data and that the product description did not include that information. Unknown condition (Unk) indicates that the product description did not specify fresh, frozen, or previously frozen.

As was discussed previously, the diversion ratio provides an alternative to the price elasticity that corrects for the market shares to get a more accurate depiction of the substitution pattern across products for a small change in price. The median diversion ratios are presented in table 5, with distributions for diversion to the outside good presented in figure 3. Following Conlon and Mortimer (2021), we present the diversion to the outside good on the diagonal, with off-diagonal elements indicating the rate of substitution to the column product from the row product as the row product price increases. These diversion ratios to the outside good in table 5 reveal that there is minimal diversion (0.05) to the outside good from the unknown condition Atlantic salmon products, with consumers switching to purchasing another salmon fillet almost 95% of the time. This type of finding is not evident from the price elasticities alone. The fresh and frozen products have a greater diversion to the outside good, ranging from 0.38 to 0.74, indicating that as these products increase in price there is more substitution to the excluded salmon products and other seafood. After adjusting the cross-price elasticities for the market shares, the fresh Atlantic salmon product still appears to be a common substitute for other fresh salmon products and for frozen Atlantic salmon.

The results can also be used to compute the aggregate elasticity for each market.[14] This measures how the demand for all included products would change with a proportional increase in the price of all these products. If the homogeneity of degree zero assumption that is typically made in the AIDS model is satisfied, then these aggregate elasticities should be zero. However, the aggregate elasticities computed for a 10% increase in price of all included products ranges across markets from –0.332 to –0.003 with an average of –0.142. With this distribution of estimates, homogeneity of degree zero does not appear to be supported by the data. Although a variant of the AIDS model can be estimated without imposing homogeneity, this severs the connection to utility

14. These market-level elasticity estimates can also be used to explore the spatial distribution of price elasticities and diversion ratios. We performed such an analysis (see online appendix figure A.1), but no clear patterns emerged.

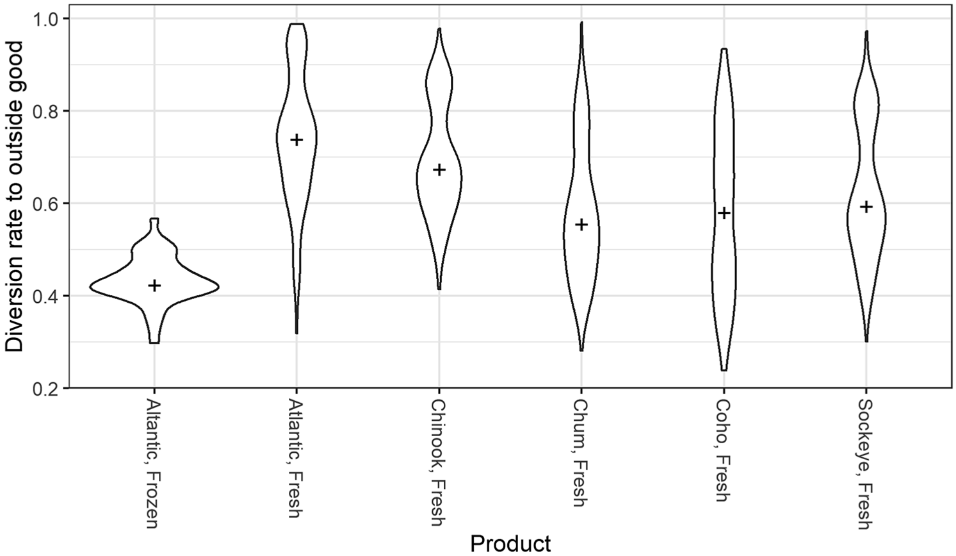

Figure 3. Distribution of Diversion Ratio to Outside Good for Selected Products

theory. This highlights that the lack of the homogeneity/weak separability assumption is another advantage of the BLP model.

CONCLUDING REMARKS

In our empirical application we use state of the art procedures, using both differentiation instruments and the approximation to the optimal instruments, to implement the random coefficients logit model and estimate the demand for salmon. The results suggest that the product condition is the most important variable for determining the substitution pattern across salmon products. With many retailers having salmon in the freezer, at the meat counter, and in the refrigerated meat section, this suggests that consumers are unlikely to turn to another section of the store to find a substitute for their first-choice product. Where reported in the data, the product conditions of fresh, frozen, and previously frozen showed that a sizable share of consumers substituted to other seafood products (likely in that section). Meanwhile, for the products with an unknown condition (possibly the tray packs in the meat section), 95% of the substitution is to other salmon fillets. Species has surprisingly little effect on the substitution pattern relative to condition, but the mean utilities are nevertheless interesting. Surprisingly, the premium species of Chinook and sockeye have negative and statistically significant coefficients relative to the baseline Atlantic salmon. However, given that Atlantic salmon are farmed while all sockeye and most Chinook are wild-caught, it is important to include the coefficient on wild-caught in comparing the species. With this inclusion, the results indicate that wild-caught sockeye and Chinook are preferred to farmed Atlantic salmon on average, although this is not statistically significant.

In addition to having a significant impact on the mean utility, product condition evidenced significant preference heterogeneity correlated with income, suggesting that higher income leads to stronger preferences for fresh salmon and lower income leads to stronger preferences for frozen salmon. The income interaction with the constant was also highly significant, indicating that higher-income households are more likely to purchase salmon fillets than other salmon and seafood products. The economic and statistical significance of the sociodemographic variable stands in contrast to many RPL studies, and is likely a consequence of observing aggregate sales data rather than individual decisions. We find that although the model allows for preference heterogeneity over price, neither the random preference nor the income interaction on price was statistically significant. This is likely due to a relatively small variation in the price across the products (by volume sold, 72.6% of products are between $5 and $10 per pound). It is interesting to note that preference heterogeneity is statistically significant for each species, but with Chinook and sockeye the preference heterogeneity is related to income, while for chum and coho it is the normally distributed unobserved heterogeneity that matters.

With the recent popularity of the random parameters logit model for analyzing individual consumer choice data to understand consumer seafood demand, our study highlights the power of the analogous random coefficients logit model for aggregate seafood sales data. Not only does this expand on the types of data that can be used for modern economic research regarding seafood demand, particularly useful as Nielsen is making their data more accessible to academic research, but it also expands on the types of questions that can be addressed. With this model it is straightforward to produce commonly presented measures such as willingness to pay estimates and price elasticities (built into most statistical packages),[15] while also enabling computation of diversion ratios. Diversion ratios provide an additional meaningful way of understanding the substitution patterns between products in a market, although they are more often used to estimate market power for merger analysis (Conlon and Mortimer 2021). Furthermore, the model streamlines the inclusion of demographic interactions compared with random parameter logit models. These demographic interactions are more frequently significant in aggregate choice data and can provide a richer understanding of the factors affecting demand. The random coefficients logit model has other advantages as well: accounting for product characteristics that are not observed by the researcher through instrumental variables that can be constructed from the data, avoiding the possible issues in product aggregation, and conforming to utility theory without making a homogeneity or weak separability assumption. The chief disadvantages are the computational complexity (our final model took over 36 hours to run) and the sensitivity to the definition of the market size (i.e., the outside good market share). Overall, we think that this discrete choice modeling approach for aggregate data offers numerous opportunities for meaningful seafood demand research.

REFERENCES

Amberg, S. M., and T. E. Hall. 2008. "Communicating Risks and Benefits of Aquaculture: A Content Analysis of US Newsprint Representations of Farmed Salmon." *Journal of the World Aquaculture Society* 39 (2): 143–57.

Ankamah-Yeboah, I., F. Asche, J. Bronnmann, M. Nielsen, and R. Nielsen. 2020. "Consumer Preference Heterogeneity and Preference Segmentation: The Case of Ecolabeled Salmon in Danish Retail Sales." *Marine Resource Economics* 35 (2): 159–76.

Asche, F. 1996. "A System Approach to the Demand for Salmon in the European Union." *Applied Economics* 28 (1): 97–101.

15. For instance, both the Stata BLP package and the PyBLP package have a command to compute elasticities for each market. We wrote additional Python code to compute the median elasticity for each product across the markets; we have posted this code to the PyBLP GitHub.

Asche, F., T. Bjørndal, and K. G. Salvanes. 1998. "The Demand for Salmon in the European Union: The Importance of Product Form and Origin." *Canadian Journal of Agricultural Economics / Revue Canadienne d'agroeconomie* 46 (1): 69–81.

Asche, F., H. Bremnes, and C. R. Wessells. 1999. "Product Aggregation, Market Integration, and Relationships between Prices: An Application to World Salmon Markets." *American Journal of Agricultural Economics* 81 (3): 568–81.

Asche, F., A. G. Guttormsen, T. Sebulonsen, and E. H. Sissener. 2005. "Competition between Farmed and Wild Salmon: The Japanese Salmon Market." *Agricultural Economics* 33 (3): 333–40.

Asche, F., T. A. Larsen, M. D. Smith, G. Sogn-Grundvåg, and J. A. Young. 2015. "Pricing of Eco-Labels with Retailer Heterogeneity." *Food Policy* 53:82–93.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.

Bjørndal, T., K. G. Salvanes, and J. H. Andreassen. 1992. "The Demand for Salmon in France: The Effects of Marketing and Structural Change." *Applied Economics* 24 (9): 1027–34.

Bronnmann, J., and F. Asche. 2017. "Sustainable Seafood from Aquaculture and Wild Fisheries: Insights from a Discrete Choice Experiment in Germany." *Ecological Economics* 142 (December): 113–19.

Carson, R. T., and M. Czajkowski. 2019. "A New Baseline Model for Estimating Willingness to Pay from Discrete Choice Models." *Journal of Environmental Economics and Management* 95 (May): 57–61.

Conlon, C., and J. Gortmaker. 2020. "Best Practices for Differentiated Products Demand Estimation with PyBLP." *RAND Journal of Economics* 51 (4): 1108–61.

Conlon, C., and J. H. Mortimer. 2021. "Empirical Properties of Diversion Ratios." *RAND Journal of Economics* 52 (4): 693–726.

Czajkowski, M., and W. Budziński. 2019. "Simulation Error in Maximum Likelihood Estimation of Discrete Choice Models." *Journal of Choice Modelling* 31 (June): 73–85.

Deaton, A., and J. Muellbauer. 1980. "An Almost Ideal Demand System." *American Economic Review* 70 (3): 312–26.

DeVoretz, D. J., and K. G. Salvanes. 1993. "Market Structure for Farmed Salmon." *American Journal of Agricultural Economics* 75 (1): 227–33.

Dube, J.-P. H., A. Hortaçsu, and J. Joo. 2020. "Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares." *SSRN Electronic Journal*, March.

FAO (Food and Agriculture Organizaiton). 2020. *The State of World Fisheries and Aquaculture 2020: Sustainability in Action*. Rome: FAO.

Foley, P., and K. Hébert. 2013. "Alternative Regimes of Transnational Environmental Certification: Governance, Marketization, and Place in Alaska's Salmon Fisheries." *Environment and Planning A: Economy and Space* 45 (11): 2734–51.

Gandhi, A., and J. F. Houde. 2019. "Measuring Substitution Patterns in Differentiated-Products Industries." *SSRN*, October.

Grebitus, C., H. H. Jensen, and J. Roosen. 2013. "US and German Consumer Preferences for Ground Beef Packaged under a Modified Atmosphere—Different Regulations, Different Behaviour?" *Food Policy* 40 (June): 109–18.

Kabir, M., and N. B. Ridler. 1984. "The Demand for Atlantic Salmon in Canada." *Canadian Journal of Agricultural Economics / Revue Canadienne d'agroeconomie* 32 (3): 560–68.

Knapp, G., C. Roheim, and J. Anderson. 2007. *The Great Salmon Run: Competition between Wild and Farmed Salmon*. Washington, DC: WWF.

Lancaster, K. J. 1966. "A New Approach to Consumer Theory." *Source: Journal of Political Economy* 74 (2): 132–57.

Lewbel, A. 1996. "Aggregation without Separability: A Generalized Composite Commodity Theorem." *American Economic Review* 86 (3): 524–43.

Mariel, P., and J. Meyerhoff. 2018. "A More Flexible Model or Simply More Effort? On the Use of Correlated Random Parameters in Applied Choice Studies." *Ecological Economics* 154 (December): 419–29.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by P. Zarembka, 105–42. Academic Press.

McFadden, D., and K. Train. 2000. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics* 15 (5): 447–70.

Nevo, A. 2000. "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand." *Journal of Economics and Management Strategy* 9 (4): 513–48.

———. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69 (2): 307–42.

NOAA Fisheries. 2019. "NOAA Fisheries Priorities and Annual Guidance 2019."

Ray, K. D., D. K. Lew, and R. Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." *Marine Resource Economics* 37 (2): 135–54.

Reynaert, M., and F. Verboven. 2014. "Improving the Performance of Random Coefficients Demand Models: The Role of Optimal Instruments." *Journal of Econometrics* 179 (1): 83–98.

Singh, K., M. M. Dey, and P. Surathkal. 2012. "Analysis of a Demand System for Unbreaded Frozen Seafood in the United States Using Store-Level Scanner Data." *Marine Resource Economics* 27 (4): 371–87.

Teisl, M. F., B. Roe, and R. L. Hicks. 2002. "Can Eco-Labels Tune a Market? Evidence from Dolphin-Safe Labeling." *Journal of Environmental Economics and Management* 43 (3): 339–59.

Villas-Boas, S. B. 2007. "Vertical Relationships between Manufacturers and Retailers: Inference with Limited Data." *Review of Economic Studies* 74 (2): 625–52.

Von Hippel, P. T., D. J. Hunter, and M. Drown. 2017. "Better Estimates from Binned Income Data: Interpolated CDFs and Mean-Matching." *Sociological Science* 4:641–55.

Wessells, C. R., and J. E. Wilen. 1993. "Economic Analysis of Japanese Household Demand for Salmon." *Journal of the World Aquaculture Society* 24 (3): 361–78.

Xie, J., H. W. Kinnucan, and Ø. Myrland. 2009. "Demand Elasticities for Farmed Salmon in World Trade." *European Review of Agricultural Economics* 36 (3): 425–45.

Xie, J., and Ø. Myrland. 2011. "Consistent Aggregation in Fish Demand: A Study of French Salmon Demand." *Marine Resource Economics* 26 (4): 267–80.