Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

## Online Appendix

*Data Cleaning*

In order to take full advantage of the data, we undertook a multi-step process to clean and trim the data, with particular effort given to inferring missing values for product attributes. We first corrected for glaring inconsistencies, then made reasonable inferences based on the observable product descriptions, trimmed observations with prices that were outliers, and finally used a series of hedonic price regressions to make inferences based on the data.

The first step of the data cleaning involved searching for inconsistencies which could be easily corrected. For instance, the county name variable included both “San Bernardino” and “San Brnrdno” and the species variable included both “Silver” and “Coho”, which are in fact two names for the same species. These inconsistencies were easily recognizable just from looking at the list of unique values for the variables. The next step was to use the 30 character product descriptions to make inferences where the data was recorded as unknown. This required an overview of the complete list of unique product descriptions for which variables were recorded as unknown.

For the production method variable, product descriptions containing “Wil”, “Wld”, and “Wc” were recorded as “Wild-caught” and product descriptions containing “far” and “frm” are recorded as farmed. For the species variable, “Alant” and “Atlant” are recorded as Atlantic, “Keta” and “Kta” are recorded as Chum, “Kg” is changed to Chinook (the proper name for King salmon), “Skeye” is recorded as Sockeye, and “Slvrbrte” is changed from Silver to Chum.

For the origin variable “Alk” was recorded as Alaska, and “C Riv” and “C Rvr” were recorded as Copper River. The process was also aided by a thorough internet search of each

Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

brand name, which revealed that Verlasso salmon is all farmed in Chile. Observations claiming to be Alaskan farmed salmon had their origin changed to "Generic" due to the regulatory impossibility of this combination. The product form variable was modified such that "Fllt" and "Fil" are recorded as Fillets, with both permutations of "H&G" which abbreviates "headed and gutted" as well as "Head On" recorded as "Whole fish". Furthermore, one product form value in the data was "tips" but internet research suggested this was not a known preparation of seafood. Representing less than one percent of all observations, this seemed suspicious. Looking at all the product descriptions for which this value was applied, they contained the phrases "Fillet" or "half fish" followed by "Tp". Additional research revealed that "Tp" is an abbreviation for "tray pack" which refers to how the product is packaged at the grocery store. Therefore these values were all changed to "Fillet" since the half fish appears to be a fillet without further trimming.

The list of product descriptions was also helpful in identifying products which did not belong in the data. This search revealed two products that had been reported as salmon did not belong in the data at all. One was a strawberry-banana nectar drink carton, and the other was a deluxe pepperoni pizza. Combined with the brand research, this step also revealed that Aquamar Seafoods does not sell salmon, but rather sells surimi-based "Salmon Flavored Seafood", which was removed from the data.

After performing these analyses of the character variables, we searched for outliers for the average retail price. Because pruning the data can introduce bias to the estimates (Andrews and Currim 2005), we were very conservative in our decisions to remove observations. On the high end, we removed observations in excess of $55 per pound. This level was chosen because there are observations of a $54.99 per pound Chinook salmon product and $49.98 per pound

Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

smoked Sockeye. Above this level, there are only five observations. Of these five observations, two of them have a product description reading "Deleted # Deleted #" for $198 per pound, a generic "Salmon salmon skinless" for $194 per pound, a Chinook salmon fillet for $146.50 per pound, and a whole pink salmon for $117.99 per pound. In light of the massive price jump and the product descriptions which do not justify these high prices, these observations were dropped. On the low end, there are nine observations of wild-caught sockeye salmon at prices below $0.15 per pound which only sold one pound in the week. Nothing about the products seems to justify the low price, and the low sales volume would not be expected if the product were actually available at this price. These observations were dropped, while the observations above $0.15 were kept, since these observations were either low value species like Chum or low value product forms like "Heads and Bones".

At this juncture, a series of frequency cross-tables were generated to check for any combinations which might provide enough information for inference. Then, where appropriate, a hedonic price regression was run to examine the differences between the products with unknown characteristics and the observed products. A decision was then made on the basis of both the economic and statistical significance of the resulting point estimates.

The cross-table of species and production method reveals that there are unknown production methods for every species in the data. For Atlantic salmon, it follows that the unknowns are most likely to be farmed, and indeed the hedonic regression fails to reject the hypothesis that the unknown production methods have the same price as farmed. Therefore, the unknown Atlantic salmon are recorded as farmed. Surprisingly though, there are 3,706 observations purporting to be wild-caught Atlantic salmon. A data query reveals that these observations belong to three different product descriptions, only one of which is clearly

Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

indicated as wild-caught Scottish salmon. The product containing "wild-caught" in the product description has an average price of $17 per pound, while the other two products with generic descriptions and unknown origins have an average price of $12 per pound. In light of these observations, only the clearly labeled one was given wild-caught status in the data, with the others re-coded to farmed. This left only 53 observations of wild-caught Atlantic salmon. For Chinook and Coho, the regression revealed that the unknown production method products had a price that did not significantly differ from wild-caught, but was significantly greater than farmed. Therefore these unknowns were changed to wild-caught. For Chum, Pink, and Sockeye salmon, there were zero observations of farmed products as they were not commercially farmed in the sample period. The regression analysis reports statistically significant differences between the unknowns and the wild-caught labeled products, but with wild-caught being only slightly lower in price in all three cases (estimates of –0.09, –0.16, and –0.04 respectively) it is reasonable to assume these were all wild-caught.

The cross-table of species and origin revealed some issues with origin for Atlantic salmon in the data, and a large share of unknown origins for Pacific salmon species. For the Atlantic salmon, the most frequent origin is "Atlantic" which provides no meaningful information to the researcher or the consumer. Some observations were changed from Alaskan to "Generic" origin in the first stage of cleaning, but some Alaskan Atlantic salmon observations remained. The regression analysis shows that unknown origin Atlantic salmon command a statistically different but only slightly higher price than those labeled in the data as Atlantic and Alaskan (estimates of –0.14 and –0.41 respectively). We chose to combine the Atlantic, Alaskan, and Unknown origins into one "Generic" origin. Chinook salmon is only recorded with origins of Alaska and Copper River, with the regression analysis showing that Copper River commands a

significant premium while Alaskan origin has a significantly lower price. Because the majority of Chinook are caught in Alaska, the negative estimate likely has more to do with unobserved qualities in the data than consumer preferences, therefore the unknown origin Chinook salmon are changed to Alaskan origin. For Chum salmon, the unknown origin products have a lower price than the Alaskan origin products, with a highly significant point estimate of –0.285. Therefore, these unknowns were defined as "Generic" origins as with Atlantic salmon. Coho salmon are reported as coming from Alaska, Chile, and Copper River, with the regression analysis suggesting significant price differences from the unknowns for all three regions. Therefore, the observations of unknown origin Coho were also changed to "Generic". Every observation of Pink salmon has an unknown origin, therefore these were simply recorded as "Generic". The Sockeye salmon are coming from Alaska and Copper River, with the regression analysis showing a significant premium for Copper River but no statistical difference with Alaskan origin. The unknown origin Sockeye salmon are changed to Alaskan Origin. Observations with the species unknown but an origin of Norway or Scotland can safely be coded to Atlantic salmon. The regression analysis was highly successful at providing guidance to where the unknowns could safely be classified, with only a few cases in which economic significance led to a decision in contradiction of the statistical significance.

The cross-table of origin and production method suggest it should be rather straightforward to assign the unknowns. Unknown production method products from Norway and Scotland are almost certainly farmed, and the regression analysis confirms that the unknowns are not statistically different in price from the farmed products. Thus, the unknowns are changed to

Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

farm-raised. Conversely, Copper River and Alaska should have no farming whatsoever due to the statewide ban on finfish farming, [1] and therefore were coded to wild-caught.

Researching the brands online revealed that some unknown values could be inferred from the brand name alone. Loch Duart, LTD exclusively sells farmed Scottish Atlantic salmon and C. Wirthy & Co. exclusively sell frozen farm-raised Atlantic salmon.

While it was a challenging and time-consuming process to clean the data, the end result had significantly fewer unknowns without having to remove a large number of observations. Many of the issues encountered may be specific to this particular data set, as scanner data do not typically contain detailed product characteristics such as country of origin or product form. Nevertheless, these methods may also be applied when using researcher-assigned characteristics on the basis of product descriptions and UPC data.

[1] https://www.adfg.alaska.gov/index.cfm?adfg=fishingaquaticfarming.main

Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

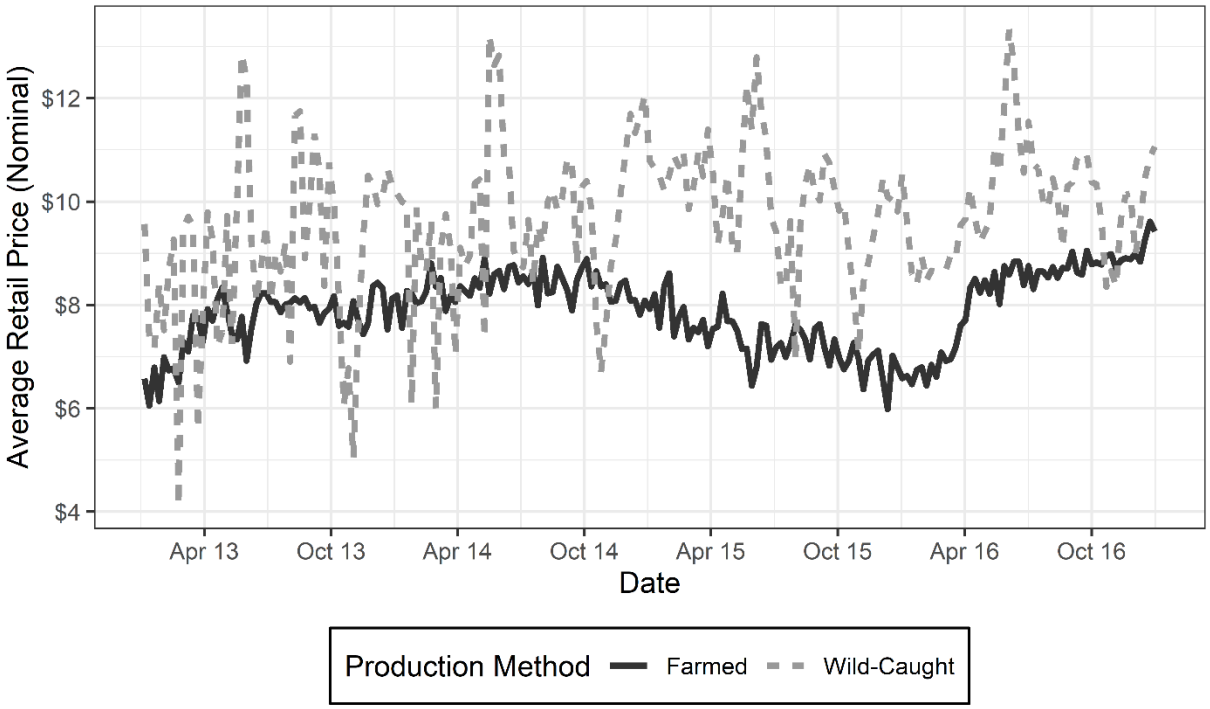


Figure A1: Weekly Price of Salmon by Production Method

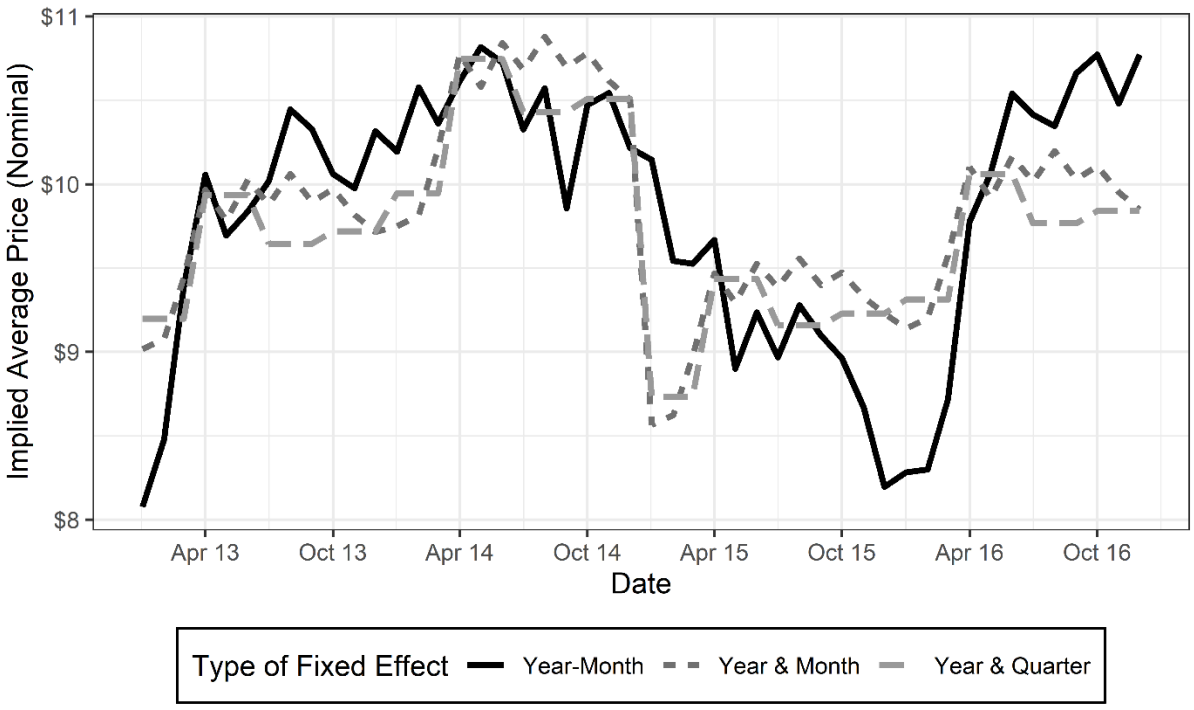


Figure A2: Comparison of implied price path for various time fixed effect strategies

Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

Table A1: Robust and Clustered Standard Errors and Significance for Pooled Model

| | Coeff. | Robust SE | Clustered SE | | | |
|---|---|---|---|---|---|---|
| | | | County | Species | Spcs x YM | Spcs x Cnty |
| Wild | 0.36 | 0.02*** | 0.02*** | 0.10*** | 0.10*** | 0.10*** |
| Hds & Bns | –1.66 | 0.02*** | 0.14*** | 0.03*** | 0.04*** | 0.06*** |
| Portion | –0.32 | 0.01*** | 0.04*** | 0.05*** | 0.05*** | 0.05*** |
| Steak | 0.01 | 0.00*** | 0.02 | 0.01 | 0.01 | 0.01 |
| Whole | –0.66 | 0.01*** | 0.04*** | 0.07*** | 0.08*** | 0.08*** |
| Unknown | –0.03 | 0.00*** | 0.03 | 0.09 | 0.09 | 0.09 |
| Fresh | –0.04 | 0.00*** | 0.04 | 0.06 | 0.06 | 0.07 |
| Frozen | –0.12 | 0.01*** | 0.03*** | 0.05* | 0.05** | 0.06** |
| Prv. Frzn | –0.07 | 0.01*** | 0.02*** | 0.05 | 0.05 | 0.06 |
| Alaska | 0.18 | 0.02*** | 0.02*** | 0.09* | 0.10* | 0.08** |
| Chile | 0.80 | 0.02*** | 0.04*** | 0.16*** | 0.16*** | 0.16*** |
| Cpr River | 0.36 | 0.02*** | 0.04*** | 0.08*** | 0.10*** | 0.08*** |
| Norway | 0.14 | 0.00*** | 0.02*** | 0.05** | 0.05*** | 0.05*** |
| Scotland | 0.44 | 0.03*** | 0.03*** | 0.05*** | 0.06*** | 0.05*** |
| Chinook | 0.08 | 0.01*** | 0.02*** | 0.09 | 0.09 | 0.08 |
| Chum | –0.64 | 0.02*** | 0.02*** | 0.12*** | 0.12*** | 0.11*** |
| Coho | –0.19 | 0.02*** | 0.02*** | 0.10* | 0.10** | 0.09** |
| Pink | –0.92 | 0.02*** | 0.02*** | 0.14*** | 0.14*** | 0.13*** |
| Sockeye | –0.24 | 0.01*** | 0.01*** | 0.11* | 0.12** | 0.11** |

Note: Critical values drawn from Student's t distribution with degrees of freedom equal to number of clusters minus one. Asterisks represent the following: *** = p<0.01, ** = p<0.05, * = p<0.1.

Supplemental Material (not copyedited or formatted) for: Kevin Daniel Ray, Daniel K. Lew, Rosemary Kosaka. 2022. "Hedonic Price Functions and Market Structure: An Analysis of Supply-Motivated Submarkets for Salmon in California." Marine Resource Economics 37(2). DOI: https://doi.org/10.1086/718479.

Table A2: Robust and Clustered Standard Errors and Significance for Atlantic Salmon Only

| | Coeff. | Robust SE | Clustered SE | | |
|---|---|---|---|---|---|
| | | | County | Year-Month | Two-way |
| Wild | –0.42 | 0.03*** | 0.08*** | 0.06*** | 0.09*** |
| | | | | | |
| Heads & Bones | –1.63 | 0.02*** | 0.17*** | 0.05*** | 0.17*** |
| Portion | –0.37 | 0.00*** | 0.03*** | 0.01*** | 0.03*** |
| Steak | 0.00 | 0.00 | 0.02 | 0.01 | 0.02 |
| Whole | –0.33 | 0.01*** | 0.03*** | 0.03*** | 0.04*** |
| Unknown | 0.09 | 0.01*** | 0.04** | 0.02*** | 0.04** |
| | | | | | |
| Fresh | –0.12 | 0.00*** | 0.04*** | 0.01*** | 0.04*** |
| Frozen | –0.55 | 0.01*** | 0.03*** | 0.04*** | 0.04*** |
| Prev. Frozen | –0.43 | 0.02*** | 0.07*** | 0.04*** | 0.07*** |
| | | | | | |
| Chile | 0.24 | 0.01*** | 0.02*** | 0.02*** | 0.03*** |
| Norway | 0.08 | 0.00*** | 0.02*** | 0.01*** | 0.03*** |
| Scotland | 0.56 | 0.02*** | 0.07*** | 0.04*** | 0.07*** |

Note: Critical values drawn from Student's t distribution with degrees of freedom equal to number of clusters minus one. Asterisks represent the following: *** = p<0.01, ** = p<0.05, * = p<0.1.

Table A3: Robust and Clustered Standard Errors and Significance for Sockeye Salmon Only

| | Coeff. | Robust SE | Clustered SE | | |
|---|---|---|---|---|---|
| | | | County | Year-Month | Two-way |
| Portion | –0.14 | 0.01*** | 0.02*** | 0.02*** | 0.025*** |
| Steak | –0.10 | 0.03*** | 0.05** | 0.08 | 0.082 |
| Whole | –0.64 | 0.06*** | 0.12*** | 0.16*** | 0.182*** |
| Unknown | –0.14 | 0.01*** | 0.01*** | 0.02*** | 0.023*** |
| | | | | | |
| Fresh | 0.10 | 0.01*** | 0.01*** | 0.03*** | 0.03*** |
| Frozen | –0.10 | 0.00*** | 0.01*** | 0.02*** | 0.02*** |
| Prev. Frozen | –0.05 | 0.00*** | 0.02** | 0.02*** | 0.03* |
| | | | | | |
| Alaska | –0.22 | 0.01*** | 0.02*** | 0.05*** | 0.05*** |

Note: Critical values drawn from Student's t distribution with degrees of freedom equal to number of clusters minus one. Asterisks represent the following: *** = p<0.01, ** = p<0.05, * = p<0.