9.0 APPENDIX 1 - Localized Depletions of Pacific Cod in the Eastern Bering Sea

Statistical Models of Catch Per Unit Effort: Localized Depletions of Pacific Cod in the Eastern Bering Sea

A PRELIMINARY DRAFT

By

Martin D. Smith
Ph.D. Candidate
Department of Agricultural and Resource Economics
University of California, Davis
Davis, CA 95616

Prepared

for

Alaska Fisheries Science Center
National Marine Fisheries Service
August 13, 2000

Statistical Models of Catch Per Unit Effort:
Localized Depletions of Pacific Cod in the Eastern Bering Sea

The goal of this report is to apply regression analysis and other statistical techniques in order to answer the question of whether the Alaskan trawl fishing fleet causes localized depletions of Pacific cod populations intra-seasonally. Fritz (1998), Smith (1998), and Smith (1999) performed similar analyses on Alaskan Atka mackerel.

There are four sections of this report. Similar to Smith (1999), Section 1 extends the closed population Leslie regression framework using Ordinary Least Squares on disaggregated Catch Per Unit Effort (CPUE) data and Generalized Least Squares on daily data. The Leslie framework is then relaxed to include another regressor, and tests for differences in conditional means at different times within a season are conducted. I also perform a Least Squares Dummy Variable Regression for one region/season. Equivalent to a Fixed Effects Model, this technique accounts for boat skipper heterogeneity. Section 2 introduces alternative definitions of fishing effort and runs the relevant regressions from section one using these alternative measures. Section 3 discusses the ability to make inferences about population migrations using spatially explicit data. Nonparametric analyses are conducted to assess whether or not CPUE declines are attributable to emigration. Section 4 presents a preliminary statistical model of CPUE using spatially explicit regressors. Though it is not a complete structural model of the fish population dynamics, it is a reduced form formalization of the nonparametric analyses conducted in Section 3. Section 5 synthesizes results and presents conclusions.

Overall, I find results that are consistent with localized depletions of Pacific cod. All of the closed population regressions have a negative and statistically significant coefficient on cumulative catch. This indicates the CPUE declines could be attributable to fishing mortality. An alternative possibility is that CPUE declines are attributable to fish dispersal or fish dispersion. Here it is important to distinguish between dispersal and dispersion as the terms are used in this report. Fish dispersal in this report means that cod migrate in mass from one location to another location without changing the density of the school. Fish dispersion, on the other hand, means that in the course of a fishing season, cod move from high density schools to lower density populations that are more uniformly distributed within a given area. If fish dispersion does not take place within a season and a closed population model adequately describes the cod population, then CPUE declines are likely due to fishing mortality. Although it is important to emphasize how strong assumptions of a closed population are, analysis of the catch data in an open population framework provides results that are consistent with the closed population conclusions. In particular, spatially explicit analyses do not strongly indicate that CPUE declines inside the Sea Lion Conservation Area are due to dispersal, i.e. fish migrations into other fishing grounds. However, the analysis conducted in this report does not eliminate the possibility that CPUE declines are due to fish dispersion. With the data and analysis available at this time, it is not possible to assign causality to fishing mortality, to dispersion, or to both factors.

Section 1

This section provides both Ordinary Least Squares (OLS) regressions of disaggregated CPUE data and Generalized Least Squares (GLS) regressions of daily aggregated data.[1] The justification for this approach is that higher levels of aggregation are not using available information efficiently. Recall that the Leslie model with a stochastic component posits the following relationship:

$$\text{(1)} \qquad \frac{C_t}{f_t} = qB_0 - qK_t + e_t$$

where C_t is current catch, f_t is effort, q is catchability, B_0 is underlying biomass, and K_t is cumulative catch, and e_t is a an error term with mean of 0 and variance σ^2.[2] One can rewrite equation (1) as the following regression equation:

$$\text{(2)} \qquad CPUE_t = \alpha + \beta K_t + e_t$$

The question of aggregation is essentially two related questions: 1) what to do when there are multiple observations at the same time, and 2) what the appropriate time step is. These questions are essentially theoretical ones about the data generating process. As a practical matter, I use one day as the minimum time step. Thus, OLS on the disaggregated data uses multiple observations for each day (one observation for each haul). Since K_t captures cumulative catch excluding contemporaneous catch, for each day K_t is the same across all $CPUE_t$.

Table 1 presents OLS results for sampled trawl hauls in 1997 and in 1998 using disaggregated data.[3] Given that a closed population model correctly describes the data

[1] The text and methodologies follow Smith (1999).
[2] Following Smith (1998), cumulative catch does not include contemporaneous catch. Thus, the problems of bias and inconsistency are avoided.
[3] Note that cumulative catch data for 1997 regressions throughout this report include cumulative catch within the SCA of the trawl fleet only. The 1998 regressions, in contrast, are run using cumulative catch for all gear types in the SCA.

generating process and that CPUE reflects the underlying biomass, Table 1 provides evidence of localized depletion. All of the coefficients on cumulative catch are negative and statistically significant.

A second approach to using a daily time step is to average the CPUE over all hauls at time t. This approach still uses all of the observations. The essential difference among daily aggregation, weekly aggregation, and no aggregation is the weight attached to each disaggregated observation. With no aggregation, extreme observations (very high or very low CPUE) receive more weight than with aggregation. On the other hand, with a large time step such as weekly aggregation, very little weight is attached to an extreme observation, which may provide a lot of information about the mean CPUE. More importantly, some time steps will contain more hauls and thus more information about CPUE. To reconcile the tradeoff between aggregated and disaggregated data, I use Generalized Least Squares (GLS) techniques to analyze daily aggregated data.[4] This approach posits a heteroskedastic error term such that the diagonal elements of the covariance matrix are a function of exogenous variables. In particular, I use two weighting schemes, the inverse of number of hauls and the inverse of number of vessels. The idea is simply that the average CPUE on a given day contains more information if there were many hauls that day than if there were few hauls. Similarly, average CPUE on a given day contains more information when there are many vessels than when there are few.

Table 2 contains results from the GLS approach for sampled hauls in 1997 and in 1998 using number of hauls as weights. All regression coefficients are statistically significant at the 5% level. As in Table 1, the signs of the coefficients are all consistent with the hypothesis of

[4] The technique I use is also known as Weighted Least Squares (WLS). The semantic distinction is whether the weighting matrix chosen is the true form of the covariance matrix. If the weighting matrix chosen is the truth, WLS is equivalent to GLS and is efficient.

depletion. Table 3 contains results from the GLS using number of vessels as weights. All regression coefficients are statistically significant at the 5% level. Again as in Tables 1 and 2, the signs of the coefficients are all consistent with the hypothesis of depletion.

A criticism of these statistical analyses of CPUE is that data near the end of the season do no reflect the true effort levels. There are two relevant issues, both of which can be addressed using statistical analysis. First, the percentage of cod caught in a haul may decrease in the end of the season because boats choose to target higher value species. I address this issue at the end of this section. Second, allegedly, nets are left in the water and not fishing when boats are filled to near capacity. My understanding of the data is that the effort variable only records hours of fishing. However, if the effort data include some observations for which nets are left idle in the water, a GLS framework is appropriate. I calculate weighted least squares regressions for several weighting schemes using the disaggregated data. Table 4 reports results for GLS using three different weighting schemes. The regressions are weighted inversely by the net time. This captures the idea that observations on long hauls may include time during which the net was left idle, and such observations provide less information about the cod population. The second set of regressions uses step function weights such that longer hauls receive less weight. The last set uses equal weights and truncates observations for which the net time exceeded three hours. The results in Table 4 are all statistically significant at the 5% level and are consistent with the hypothesis of depletion.

A more general framework to consider is that CPUE does not just depend on the level of biomass; instead, CPUE depends also on characteristics of the haul. One possibility is that the by-catch level indicates the extent to which cod were targeted. A higher percentage cod, *ceteris paribus*, would lead to a higher CPUE. A regression equation, hence, ought to condition on percent cod and not just cumulative catch. Then the policy question becomes: does the conditional mean of CPUE decline as cumulative catch increases? A way to test this is simply to look at whether the forecast CPUE (for the early period) conditional on a vector of regressors $\mathbf{x_1}$ is greater than the forecast CPUE (for the later period) conditional on a vector of regressors $\mathbf{x_2}$. This yields the following general form for a t test:

$$(5) \qquad t = \frac{\mathbf{x}_1'\hat{\beta} - \mathbf{x}_2'\hat{\beta}}{\sqrt{Var(\mathbf{x}_1'\hat{\beta} - \mathbf{x}_2'\hat{\beta})}}$$

where $\hat{\beta}$ is a vector of estimated coefficients.

Suppose one fixes the percent cod as the same for the first half and the second half of the season. Then $[\mathbf{x_1}' - \mathbf{x_2}']$ reduces to the (1 x 3) vector

$[0 \ \ (K_1\text{-}K_2)\hat{\beta}_K \ \ 0]$, where K_1 and K_2 are average cumulative catch for the first and second halves respectively and $\hat{\beta}_K$ is the estimated coefficient on cumulative catch. Because the first and third elements of $[\mathbf{x_1}' - \mathbf{x_2}']$ are zero, the test statistic reduces to the t ratio for the cumulative catch coefficient in the regression of CPUE on a constant, cumulative catch, and percent cod. Results of these conditional mean tests appear in Table 5. Table 6 performs the same test but also conditions on vessel length. The test statistics in both tables are significant. This suggests that even after conditioning on the percentage cod caught and vessel length, increasing cumulative catch reduces CPUE. Again assuming that CPUE is an indicator of biomass, catch appears to reduce the biomass over the course of the season.

A Fixed Effects estimator also fits within the framework outlined above. In essence, we condition on vessel specific effects rather than just the cumulative catch. Unlike the standard setting for Fixed Effects, here we do not have complete panels of data; not all vessels in the sample are active every day in the season within a given region. Fortunately, the number of vessels is small, so we can estimate a Least Squares Dummy Variable Regression. Mathematically, this is equivalent to the Fixed Effects Estimator. The motivation for Fixed Effects in this setting is that some boats may be better fishing vessels than others. This could result from heterogeneity in vessel characteristics, gear, and/or skipper abilities.

Table 7 reports Least Squares Dummy Variable Regressions for the trawl fleet in 1997 and in 1998. Vessel dummy variables account for boat skipper heterogeneity. Though not all of the dummy variables are statistically significant, the coefficient on lagged cumulative catch is negative and significant in both regressions. This indicates that, conditional on individual boat effects, there is a negative and statistically significant relationship between cumulative catch and CPUE. Table 8 reports similar results but includes the percent cod variable. As in Table 7, the coefficient on lagged cumulative catch is negative and significant in both regressions, indicating a negative relationship between cumulative catch and CPUE.

Section 2

The idea that CPUE may depend on vessel length and percent catch in cod raises the question of whether nominal net time is the most appropriate measure of fishing effort. A measure of effective effort might be more appropriate in a fishery that has multiple species targeting strategies. For example, suppose a vessel makes a three hour trawl and catches fifty percent Pacific cod and fifty percent Atka mackerel. To assign three hours of effort to cod may overstate the fishing effort because, in fact, the vessel was targeting both cod and mackerel. Over the course of a season, targeting strategies could change. Hence, the following question arises: are CPUE declines attributable to changes in targeting strategies rather than fishing mortality? To investigate this question, I propose two definitions of effective fishing effort. Using these definitions in the remainder of this section, I find that targeting strategy changes do not account for CPUE declines.

There are two necessary adjustments to raw effort: one to account for vessel size and one to account for targeting strategy. Denote effective effort as f_t^*. For vessel size (v), I assume the following: $\frac{\partial f^*}{\partial v} > 0$ and $\frac{\partial^2 f^*}{\partial v^2} < 0$. These assumptions capture the idea that effective effort is monotonically increasing in vessel size but at a decreasing rate. A convenient functional form is to multiply effort by the natural logarithm of vessel size. Furthermore, I assume that effective effort is a linear function of a targeting coefficient (ρ) as follows:

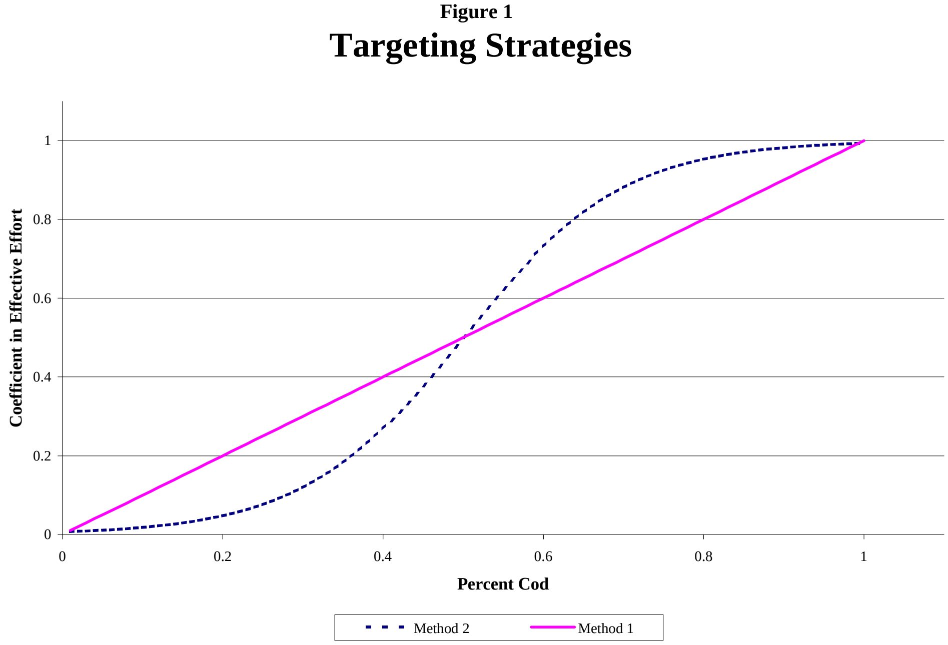
(6) $$f_t^* = \rho f_t \ln(v)$$

The simplest way to compute effective effort then is to assume that ρ is the percent cod in the haul. I refer to this approach as Effective Effort Method 1. An alternative definition for ρ recognizes that very low percent cod may indicate that cod was non-targeted by-catch. Similarly, very high percent cod may indicate that the haul was a single species target and the other species in the haul are by-catch. This suggests an inverse logistic relationship between percent cod (p) and ρ as follows:

$$\rho = \operatorname{logit}^{-1}[(p-a)b] = \frac{e^{(p-a)b}}{1+e^{(p-a)b}} \qquad (7)$$

where a is a centering parameter and b is a scale parameter. For the analysis below, I have assumed a=.5 and b=10. I refer to this approach as Effective Effort Method 2. Figure 1 depicts

Figure 1



these methods graphically.

Using Method 1 and Method 2 to calculate CPUE based on effective effort, I ran all of the regressions in Tables 1-4 for the 1998 data (12 total regressions). In all regressions, the coefficient on cumulative catch was negative and statistically significant at the 1% level. These results are qualitatively the same as the results in Section 1. Hence, after accounting for vessel length and targeting strategy, there is still a negative relationship between cumulative catch and CPUE.

Section 3

A major criticism of the Leslie framework for assessing depletion is that it assumes a closed population. A more general model would allow for natural fish mortality, emigration, and immigration within a season. Shimada and Kimura (1994) perform a tagging study that suggests migration patterns are important determinants of seasonal cod abundance in different locations. Smith (1999) applied a model following Hilborn and Walters (1992) to Atka mackerel data. This approach required several extreme assumptions about unobservables and still produced many results that were unrealistic (e.g. negative population estimates). In this section, I apply a nonparametric approach that is less structural and requires fewer assumptions to assess whether CPUE declines are attributable to fish migrations, i.e. dispersal.

The basic idea is that CPUE inside the Sea Lion Conservation Area (SCA) is a function of biomass in the SCA and net migration. When there is net emigration, biomass decreases in the SCA, leading to a CPUE decline. However, biomass must increase in other areas due to migration. This increase should be reflected in CPUE gains in these other areas. I suppose that Pacific cod in Alaska can be described by a metapopulation with three zones: Eastern Bering Sea inside the Sea Lion Conservation Area (EBS in SCA), Eastern Bering Sea outside the Sea Lion Conservation Area (EBS outside SCA), and the Gulf of Alaska (GAO). Within each area, there are three gear types involved in cod fishing: trawl, pot, and longline. Thus, when CPUE declines within the Eastern Bering Sea (EBS) in SCA, if these declines are truly due to mass emigrations, we should observe CPUE increases in Gulf of Alaska and in EBS outside the SCA.

The nonparametric analyses appear in Tables 9-14. Each cell reports a count of number of occasions on which CPUE increased or decreased in the SCA and the corresponding increase or decrease in regions outside the SCA. Each box on each table is a 2X2 matrix that compares

CPUE changes in the fishery inside the SCA with CPUE changes for one of the gear types outside the SCA (either EBS outside SCA or GOA). Recall that CPUE declines due to migration out of the SCA should be reflected in CPUE increases in other areas. Similarly, CPUE increases due to migration into the SCA should be reflected by CPUE declines in other areas. Thus, if changes in CPUE are due to migration, we should not see numbers along the diagonals of these boxes. Or at the very least, we should see smaller numbers along the diagonals than we see on the off-diagonals.

Tables 9-11 use a daily time step.[5] Table 9 reports results for the trawl fishery inside the SCA, Table 10 contains results for the pot fishery inside the SCA, and Table 11 contains results for the longline fishery inside the SCA. In all three tables, there are sizable numbers along the diagonals. These patterns are especially noteworthy in comparisons of the trawl and longline fisheries within the SCA to trawl and longline fisheries in EBS outside the SCA.

One critique of Tables 9-11 is that a daily time step is simply too short to observe spatial interdependencies in CPUEs. As an alternative, I do the analysis using a ten-day time step. CPUEs here are ten-day weighted averages. This greatly reduces the number of possible comparisons, but it provides a more plausible amount of time for fish migration signals to appear in the data. Tables 12-14 present the nonparametric results of the ten-day aggregation. Compared to the daily time step, we see even higher relative diagonal to off-diagonal numbers in these analyses. In fact, the diagonals seem to dominate the off-diagonals. Moreover, simultaneous decreases in CPUE are common throughout the tables. Of the sixteen comparisons in Table 12, simultaneous CPUE decreases are the most common event in five cases and tied for the most common event in three cases. Of the sixteen comparisons in Table 13, simultaneous

CPUE decreases are the most common event in nine cases and tied for the most common event in one case. Of the fifteen comparisons in Table 14, simultaneous CPUE decreases are the most common event in six cases.

Overall, the nonparametric results cast doubt on the idea that CPUE declines are due to large fish migrations. However, they do not indicate one way or the other whether CPUE declines are attributable to fishing mortality or to fish dispersion. Both explanations are consistent with the nonparametric findings.

[5] Note that the number of total occasions differs by gear type and effort calculation method due the presence of different amounts of data for different time periods across different gear type and location combinations.

Section 4

Though the nonparametric results in Section 3 do not support the idea that CPUE declines are attributable to fish migrations in mass, this section explores the issue further using regression. The approach is reduced form in the sense that there is not a specific structural model of spatial population dynamics underlying the regression specifications. The basic idea is the same as the idea in Section 3. If fish dispersal causes CPUE declines in the SCA, then CPUEs in other regions should be negatively correlated with CPUE in the SCA. Thus, I regress CPUE in the SCA for each gear type and effort definition (9 regressions total) on each CPUE outside the SCA for each gear type. I also include cumulative catch as a regressor and a series of dummy variables to deal with missing observations.[6] If dispersal patterns explain CPUE declines, then they should show up as negative and significant coefficients on the CPUE regressors.

Tables 15-17 report the results of these spatially linked regressions. The dependent variables are CPUE of cod inside the SCA for trawl, pot, and longline respectively. CPUE regressors all begin with “CP.” CP_1_O indicates trawl outside the SCA in EBS, CP_6_O indicates pot outside the SCA in EBS, and CP_8_O indicates longline outside the SCA in EBS. Similarly, CP_1_G, CP_6_G, and CP_8_G indicate respectively the same gear types in the Gulf of Alaska. Variables that begin with “D” are the dummy variables used to control for missing observations, i.e. observations for which the dependent variable is available but one of the independent variables is not available.

The most important result in Tables 15-17 is that none of the CPUE regressors are negative and statistically significant in any of the nine regressions. As such, when CPUE declines in the SCA, it is not attributable to CPUE increases in areas outside the SCA. This means that fish dispersal is not a likely explanation for CPUE declines. Moreover, all of the cumulative catch variables that are statistically significant have a negative sign (two regressions in Table 15 and one regression in Table 16). Hence, when controlling for spatial interconnectedness of the fish populations, cumulative fishing mortality appears to decrease CPUE.

There are several caveats relating to this analysis. First, the specification is *ad hoc* and not derived directly from an underlying structural model in the way that Leslie regression is derived. This casts some doubt on the ability to draw conclusions about what is ultimately of interest, the dynamics of the unobserved cod population. Second, a daily time step was used for these regressions. A longer time step might lead to different results. Alternatively, it might be appropriate to analyze spatial dynamics with temporal lags to allow adequate time for fish migration signals to be transmitted through fishing data. Third, some of the cumulative catch coefficients are not negative and statistically significant. This could mean that fishing mortality effects are not well measured.

[6] For some regressions, one or more regressors were dropped due to perfect multicollinearity.

Section 5

Analysis in Section 1 suggests that results of the closed population model are quite robust to different levels of aggregation and different specifications in the error structure. The robustness of these results is further confirmed by the conditional means testing that conditions on percent cod caught and vessel fixed effects. In all regressions, coefficients on lagged cumulative catch are negative and statistically significant. If cod populations are closed then CPUE declines are attributable to either fishing mortality or dispersion. Moreover, if cod populations are open and net migration equals natural mortality, CPUE declines again are attributable to either fishing mortality or dispersion. The results in Section 2 reinforce these conclusions. Sections 3 and 4 investigate whether CPUE declines are attributable to fish dispersal, i.e. migrations in mass, rather than fishing mortality or dispersion to lower densities. Both sections conclude that dispersal does not explain CPUE declines. So, we are left once more with the attribution of CPUE declines to fishing mortality, to fish dispersion, or to both factors.

To distinguish between fishing mortality and dispersion poses difficult problems. Ideally, the analyst would have long panels of spatially explicit data (multiple years) that includes natural experiments, i.e. mid-season closures of varying lengths at varying times. This would allow an analyst to build a structural model of cod population dynamics that is truly bioeconomic. Without the natural experiments, even having long panels would make it difficult to distinguish between localized effects of mortality and dispersion.

Further analysis of the existing data would include tests for structural change. In essence, one would ask whether there is a period over which CPUE does not decline in cumulative catch at the beginning of a season. If so, there is some evidence that declines later in the season are due to dispersion. However, this ignores the fact that cumulative catch and CPUE might be related nonlinearly. A thorough analysis of structural change would include alternative functional forms relating cumulative catch to CPUE.

Sources

Fritz, Lowell W., "Do Trawl Fisheries off Alaska Create Localized Depletions of Atka mackerel (Pleurogrammus monopterygius)?" Alaska Fisheries Science Center, National Marine Fisheries Service, 1998.

Hilborn, Ray and Carl J. Walters, *Quantitative Fisheries Stock Assessments: Choice, Dynamics and Uncertainty* (New York: Chapman and Halley, 1992).

Shimada, Allen M. and Daniel K. Kimura, "Seasonal movements of Pacific cod, *Gadus macrocephalus*, in the eastern Bering Sea and adjacent waters based on tag-recapture data," *Fishery Bulletin* 92: 800-816.

Smith, Martin D., "Report on Leslie Regressions of Alaskan Atka Mackerel Populations," Prepared for Alaskan Fisheries Science Center, National Marine Fisheries Service, 1998.

Smith, Martin D., "Statistical Models of Catch Per Unit Effort: An Analysis of Localized Depletions in Alaskan Atka Mackerel," Prepared for Alaskan Fisheries Science Center, National Marine Fisheries Service, 1999.