Fisheries Centre

The University of British Columbia

Working Paper Series

Working paper # 2006-22

Bayesian Estimation of Technical Efficiency in the Pacific Hake Fishery

David Tomberlin, Xavier Irz and Garth Holloway

Year: 2006

email: david.tomberlin@noaa.gov

Bayesian estimation of technical efficiency in the Pacific hake fishery[1]

David Tomberlin
National Marine Fisheries Service,
Santa Cruz, CA, USA; Email: david.tomberlin@noaa.gov

Xavier Irz
University of Reading,
Reading, England, UK; Email: x.t.irz@reading.ac.uk

Garth Holloway
University of Reading,
Reading, England, UK; Email: garth.holloway@reading.ac.uk

Abstract

Technical efficiency in fisheries has most often been investigated using data envelopment analysis or stochastic production frontiers, the latter usually estimated by maximum likelihood methods. This paper describes an alternative, Bayesian approach to estimating stochastic production frontiers in fisheries. Through an application to the US West Coast hake (also known as whiting) fishery, we demonstrate the relevance of Bayesian methods to technical efficiency analysis and highlight some particular strengths of the approach that may appeal to fisheries economists. We focus on variation in efficiency among boats and over time, showing how the Bayesian hierarchical approach links unobserved heterogeneity among sample units within a common framework. Although we are not aware of previous applications of hierarchical models to fisheries efficiency analysis, the models introduced here apply techniques common in other areas of economics, namely Markov Chain Monte Carlo (MCMC) methods. Our chief conclusions are that 1) panel models with hierarchical structure to allow for boat- and year-specific efficiency measures are preferable to simpler specifications, and 2) there appears to have been a progressive outward shift in the efficient frontier in the shore-based whiting fishery during 1987-2003.

Introduction

Analysis of technical efficiency in fisheries provides a framework for studying issues such as regulatory impacts and the relative importance of skill versus luck in fishing success[2]. The two main approaches to such analysis are data envelopment analysis (see, e.g., Walden et al., 2003), and stochastic production models (see the literature reviewed in Holloway et al. (2005)). In this paper, we investigate technical efficiency in the US Pacific hake fishery, by volume the largest fishery on the US West Coast[3], within the framework of a stochastic production frontier model.

---

[1] This paper should have been included In Sumaila, U.R., Marsden, D. (eds.), 2006. Proceedings of the 2005 North American Association of Fisheries Economists Forum FCRR 2006, 14(1) 220 pp.

[2] Overviews of technical efficiency and productivity in fisheries are available in Alvarez (2001), Felthoven and Paul (2004), and Fernandez et al. (2002).

[3] West Coast here refers to California, Oregon, and Washington.

Our emphasis is on developing and comparing models that permit changes over time in firms' locations with respect to an efficient frontier and in the location of the frontier itself. Developing these models within a hierarchical Bayesian framework also permits us to explore the possibility that year- or boat-specific efficiency measures are linked stochastically.

While most stochastic production frontiers, both within fisheries and within the wider production economics literature, are estimated using maximum likelihood techniques, we adopt the Bayesian approach here for several reasons. First, recent advances in Bayesian estimation permit implementation of models of significantly greater complexity than previously thought possible. Second, the hierarchical approach employed in our Bayesian estimation enables us to incorporate, formally and robustly, aspects of intra-sample heterogeneity that may significantly affect inference about efficiency. Third, because many questions about appropriate specification in the presence of heterogeneity reduce to a question of model selection, our Bayesian MCMC estimation has particular appeal because it leads naturally to the necessary diagnostics for model selection.

Pacific hake (*Merluccius productus*), also known as Pacific whiting, is a migratory species found from Baja California to the Gulf of Alaska. Most of the commercial catch is taken with mid-water trawl gear in the northern half of this range. The hake fishery is a high-volume, low-margin fishery that accounted for 64% of the catch (by weight) in the shore-based limited-entry trawl fleet during 1993-2003. Because of a tendency to spoil rapidly, hake were long regarded as undesirable by US fishermen. During the 1990s, growth in the surimi market and improved processing techniques led to huge increases in the domestic catch, with most boats having refrigerated seawater storage to reduce spoilage, making them a fairly distinct segment.

The hake fishery is managed by the Pacific Fisheries Management Council, which allocates the total allowable catch among a tribal fishery and three non-tribal sectors (a shore-based fleet, a catcher-processor fleet, and a factory trawler fleet). In this paper, we consider only the shore-based fleet, specifically the 41 boats that caught hake on at least 100 trips during our study period (1987-2003). While trip limits apply before and after the main season (to allow for incidental catch of whiting), during the main season there are no trip limits.

Pacific whiting were declared overfished in 2002—meaning that estimated biomass had fallen below 25% of estimated unfished biomass—but have rebounded recently. In 2004, the National Marine Fisheries Service declared the stock rebuilt, which means that current biomass is over 40% of unfished biomass. Although there are by-catch concerns (salmon and rockfish, in particular), trawlers operating off the coastal shelf can often net hake with little or no by-catch. For this reason, we adopt a single-output stochastic production function in our analysis below.

COMPETING STOCHASTIC FRONTIER MODELS OF TECHNICAL EFFICIENCY

The basis of our analysis is the composed-error model, first formalized by Aigner, Lovell and Schmidt (1977)[4]. This approach allows us to investigate catch as a function of inputs to production, random error, and inefficiency. Given data on output (here, whiting catch per trip) and regressors thought to influence output (inputs to production as well as exogenous factors), a stochastic production frontier may be specified as

(1) $$y_i = \boldsymbol{x_i}'\boldsymbol{\beta} - z_i + u_i$$

where y_i is the catch by unit i; $\boldsymbol{x_i}$ is a vector of explanatory variables; $\boldsymbol{\beta}$ is a vector of parameters that, together with $\boldsymbol{x_i}$, defines the frontier; u_i is a normally distributed sampling error term; and z_i a term denoting the distance of y_i from the frontier, i.e., z_i is a measure of the inefficiency of unit i,

[4] A review of recent developments in composed-error modeling is found in Murillo-Zamorano (2004).

which we will assume follows a truncated normal distribution. These are standard assumptions in the literature (see Kumbhakar and Lovell (2000) for a book-length treatment).

Using this basic framework, we estimate six alternative specifications of the composed-error model of technical efficiency. The purposes of this study are to explore the relative importance of regressor variables, to demonstrate the application of hierarchical modeling in the fisheries context, and to assess the effect of panel model specification on estimates of technical efficiency. Below, we report on the estimation of models that are elaborations of (1), assuming throughout a Cobb-Douglas production function. The inputs to production we consider are horsepower, crew size, and the total duration of tows on each trip. We also include an estimate of exploitable biomass in one model. Before detailing these model specifications, however, we briefly describe our study data.

Data

Our primary data source is the logbook information required of all vessels with federal limited-entry groundfish permits. Logbooks record information on each trip and tow, including species and estimated catch weight, gear used, location of fishing, and duration of tow. Supplementary data on boat characteristics were obtained from the Pacific Fisheries Information Network. In order to keep the analysis focused on the boats that form the core of the whiting fleet, we did not include in our study boats that had fewer than 100 whiting trips during 1987-2003. We further limited attention to trips made during May-September, when trip limits are not in effect. This left us with a data set consisting of 10,865 whiting trips taken by an unbalanced panel of 41 boats. These data account for 82% of the shore-based fleet's whiting catch during 1987-2003. In addition to logbook data, we used estimates of the Pacific hake population (spawning biomass) developed by the Pacific Fisheries Management Council's Groundfish Management Team (Helser et al., 2004).

Cross-Sectional Models 1 and 2

We first estimate two cross-sectional models to provide a baseline for comparison with subsequent panel models. Both models include crew size, horsepower, and total duration of tows per trip as regressors. Model 1 includes a point estimate of coast-wide exploitable biomass by year, while Model 2 includes year dummies in lieu of the biomass estimate.

Because our unit of observation is the individual trip, we elaborate equation (1) to account for three dimensions of the sampling environment, namely the vessel, year, and trip for which catch is reported. We denote catch (in pounds) by vessel i in year j during trip k as y_{ijk}. Our goal is to explain variations in catch as a function of a vector of covariates $\boldsymbol{x}_{ijk}$, an implicit measure of technical efficiency z_{ijk}, and a random error term u_{ijk}. The cross-sectional regression equation in Models 1 and 2 is then:

$$(2) \qquad y_{ijk} = \boldsymbol{x_{ijk}}'\boldsymbol{\beta} - z_{ijk} + u_{ijk}.$$

In these cross-sectional models, equation (2) is estimated directly on the trip observations, with no consideration of possible links among observations for each boat or year. That is, Models 1 and 2 are distinguished from the panel models to follow by the assumption that each z_{ijk} is independently and identically distributed from the truncated-Normal distribution, $f^{\mathrm{TN}}(z_{ijk}|\mu,\omega^2)$. We also assume that the sampling error term u is i.i.d. from a Normal distribution with mean zero and variance σ^2, or $f^{\mathrm{N}}(u_{ijk}|0,\sigma^2)$.

Further assuming a diffuse prior probability distribution for the model parameters and a Normal likelihood function for the data given model parameters enables us to proceed with our Bayesian estimation. Specifically, we derive fully conditional posterior distributions for the parameter set $\{\boldsymbol{\beta}, \sigma, \mu, \omega\}$, as detailed in Holloway et al. (2005). Given the fully conditional distributions of the parameters and latent inefficiency terms z, we use a Gibbs sampling approach to conduct inference on the posterior distribution.

Estimation results for Models 1 and 2 are given in Table 1. While we delay discussion of the results until all the models are presented, it is worth noting here that Model 2 is a direct response to the counter-intuitive finding that the biomass coefficient in Model 1 was negative and strongly significant.

Panel Models 3 and 4

The cross-sectional models presented above do not account for the likelihood that boats have something like a characteristic level of inefficiency, as well as the possibility that years may also have such a characteristic level of inefficiency, due for example to ocean conditions. In order to accomodate this likely feature of the data, we next estimate two single-layer hierarchical panel models. The idea of the hierarchical modeling structure is that observations within a sample may be probabilistically related at the unit or sub-unit level and that these relationships may be captured by representing model parameters themselves as draws from a distribution. For example, we may suppose that boats' inefficiencies in a given year are drawn from a common truncated-normal distribution, the mean of which varies over time, but that this mean in a given year is related to that of other years by virtue of coming from the same probability distribution function. In such a case, the econometric task would be to estimate both the year-specific mean inefficiencies and the parameters of whatever distribution is assumed to generate those mean inefficiencies.

Our Models 3 and 4 exploit the availability of panel data by invoking hierarchy in the parameters governing inefficiency. Specifically, in Model 3 we suppose that each boat i has a particular level of inefficiency, z_i, which is the same in all years, and that z_i is a realization from a truncated-Normal distribution $f^{\mathrm{TN}}(z_i|\mu,\omega^2)$. We then estimate the regression equation

$$y_{ijk} = \boldsymbol{x_{ijk}}'\beta - z_i + u_{ijk}. \tag{3}$$

Here we require estimates of z_i for each of the i=41 boats, as well as of the parameters μ, ω^2, and σ. Model 4 represents an analogous formulation in which the organizational unit is the year. That is, each of the j=17 years is assumed to have a particular level of inefficiency, z_j, that is the same for all boats and trips in that year, and the z_j are realizations from a common truncated-Normal distribution $f^{\mathrm{TN}}(z_j|\mu,\omega^2)$.

We did not expect Models 3 and 4 to perform particularly well, as the existing literature on efficiency strongly suggests that generally there is significant heterogeneity among firms' inefficiency levels, and heterogeneity over time is likely as well. However, we present these models here as an intermediate step in the development of Models 5 and 6 below, in which an inefficiency level is estimated for each boat in each year.

Panel Models 5 and 6

Models 3 and 4 are unrealistically constrained because they impose the assumption that (respectively) either each boat has an inefficiency level that is invariant over time or that all boats have the same inefficiency level within a given year. To address this shortcoming, Models 5 and 6 incorporate a two-layer hierarchical structure for the inefficiency terms. That is, we suppose that trips by boat i in year j have a particular efficiency level z_{ij}, but that there are both boat-specific and year-specific influences on the z_{ij}, which may be captured in hierarchical parameters to be estimated simultaneously.

Model 5 proceeds from the hypothesis that inter-boat differences are more important determinants of efficiency than are inter-annual differences. This hypothesis is embodied in a two-layer hierarchy of inefficiency parameters: the boat-year inefficiencies, z_{ij}, are drawn from a distribution $f^{TN}(z_{ij}|\mu_i,\omega^2)$ with a mean μ_i specific to each boat, and these μ_i are in turn drawn from a distribution $f^{TN}(\mu_i\,|\lambda,\theta)$. The regression equation becomes

$$(4) \qquad y_{ijk} = \boldsymbol{x_{ijk}}'\beta - z_{ij} + u_{ijk}.$$

This specification has considerably more estimands than the single-layer hierarchy of Models 3 and 4: 214 z_{ij} (one for each boat-year combination), 41 μ_i (one for each boat), plus $\omega, \lambda, \theta, \beta$ and σ.

Model 6 reverses the hierarchy of the boat and year influences, i.e., proceeds from the hypothesis that inter-annual differences in fishing conditions (perhaps catchability or ocean regime) are more important determinants of efficiency than are inter-boat differences. This hypothesis is also embodied in a two-layer hierarchy of inefficiency parameters: the z_{ij} are drawn from a distribution with a time-varying mean, $f^{TN}(z_{ij}|\mu_j,\omega^2)$, and these μ_j are in turn drawn from a distribution of year-specific mean inefficiencies $f^{TN}(\mu_j\,|\lambda,\theta)$. Here we must estimate 214 z_{ij}, 17 μ_j, plus $\omega, \lambda, \theta, \beta$ and σ.

Model Comparison

Bayesian model comparison is discussed in detail in Zellner (1996, pp. 291-318). However, until as recently as Chib (1995), model comparisons have proved intractable for all but the simplest of models. Chib (1995) develops a robust procedure, exploiting the MCMC estimation, for comparing models, and we implement that procedure here to ascertain the relative likelihoods that each of the respective models is more likely to have generated the data. Details on implementation of the marginal likelihood calculation in the context of stochastic production frontiers are given in Holloway et al. (2005).

RESULTS

Table 1 presents the basic results of the estimations performed. Due to space constraints, we do not report intercept or dummy variable parameter estimates; z_{ijk} from Models 1 and 2; z_i from Model 2; z_j from Model 3; or $\mu_i, \mu_j,$ and z_{ij} from Models 5 and 6.

All the models suggest that the output elasticities of horsepower and tow hours are significantly positive, and the magnitude of these parameter estimates themselves is quite similar across the models. There is less support overall for a significant positive elasticity of crew size, and the coefficient of biomass that we estimated in Model 1 turned out to be significant and negative. The reason for this is that biomass was declining quite steadily during the study period, thus its inclusion in the regression model serves as a time index during a period in which catch per trip was steadily increasing[5].

Comparing marginal likelihoods across models suggests that Models 5 and 6 are best supported by the data. Indeed, these two models, which differ only in that the hierarchy relating boat-level to year-level information is reversed, are very similar in terms of both parameter estimates and marginal likelihood values. The main conclusion to be drawn here is that the two-layer hierarchical panel models are preferred, and that the order of the hierarchy makes little difference in this case.

Mean percent efficiency scores vary among the models, with Model 5 estimating a significantly lower overall level of efficiency in the sample than any of the other models. Given the similarity between the preferred Models 5 and 6 in most respects, it is surprising that they differ to this degree in their estimates of mean efficiency scores, something we are looking into further.

---

[5] Thanks to Dan Huppert for helping us interpret this result.

Table 1. Main estimation results for six stochastic frontier models of technical efficiency in the Pacific hake fishery.

Estimand	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
β_1 (Horsepower)[1]	0.27 (14.77)[2]	0.33 (18.06)	0.28 (7.47)	0.35 (19.45)	0.37 (4.45)	0.49 (7.62)
β_2 (Tow hours)	0.05 (5.79)	0.07 (8.55)	0.06 (7.61)	0.07 (8.45)	0.05 (5.56)	0.05 (5.56)
β_3 (Crew size)	0.41 (6.38)	0.26 (3.98)	0.06 (0.78)	0.29 (4.55)	-0.09 (-1.06)	-0.08 (-0.99)
β_4 (Biomass)	-0.47 (-28.41)	--	--	--	--	--
σ (Sampling Error)	0.56 (12.48)	0.58 (109.68)	0.56 (146.50)	0.59 (147.35)	0.52 (144.73)	0.52 (143.72)
μ (Mean of z_{ijk}, z_i, or z_j)	1.52 (5.69)	0.46 (3.46)	0.49 (5.94)	0.19 (1.97)	--	--
ω (St.dev.of z_{ijk}, z_i, or z_j)	0.19 (1.87)	0.07 (3.09)	0.26 (8.30)	0.07 (2.27)	0.30 (14.21)	0.38 (18.37)
λ (Mean of μ_i or μ_j)	--	--	--	--	2.74 (4.36)	1.60 (6.79)
θ (St. dev. of μ_i or μ_j)	--	--	--	--	0.22 (6.22)	0.32 (2.38)
Log Marginal Likelihood[3]	-9373.96	-9731.66	-9285.59	-9848.04	-8387.14	-8442.29
Maximized Log Likelihood	-9189.07	-9532.95	-9078.43	-9673.77	-8338.03	-8314.63
Numerical Standard Error	0.0048	0.0021	0.0001	0.0077	0.0016	0.0038
Mean % Efficiency Score	0.88 (0.02)	0.96 (0.01)	0.96 (0.01)	0.98 (0.01)	0.81 (0.04)	0.88 (0.01)

[1] β terms are elasticities (a Cobb-Douglas production function is assumed).
[2] Numbers in parentheses are asymptotic t-values, except for in the last row, where they represent the standard deviation of Gibbs samples of the mean efficiency score.
[3] The log marginal likelihood of the model, calculated as in Chib (1995).

Figure 1 shows the change in the coefficients of year dummy variables over time, as estimated in Model 5. There seems to be a reasonably strong upward trend in the value of these parameters, a trend that was also evident in the results of Models 2, 3, 4, and 6. These dummy coefficients may indicate a significant outward shift of the efficient frontier due to technological change, though our approach here does not allow us to separate the effects of technological change from other factors that may shift the efficient frontier (such as ocean conditions) or, indeed, from other unobserved factors that may be influencing the dummy coefficient estimates. At a minimum, Figure 1 suggests the possibility of an interesting study of total factor productivity in the Pacific hake fishery.

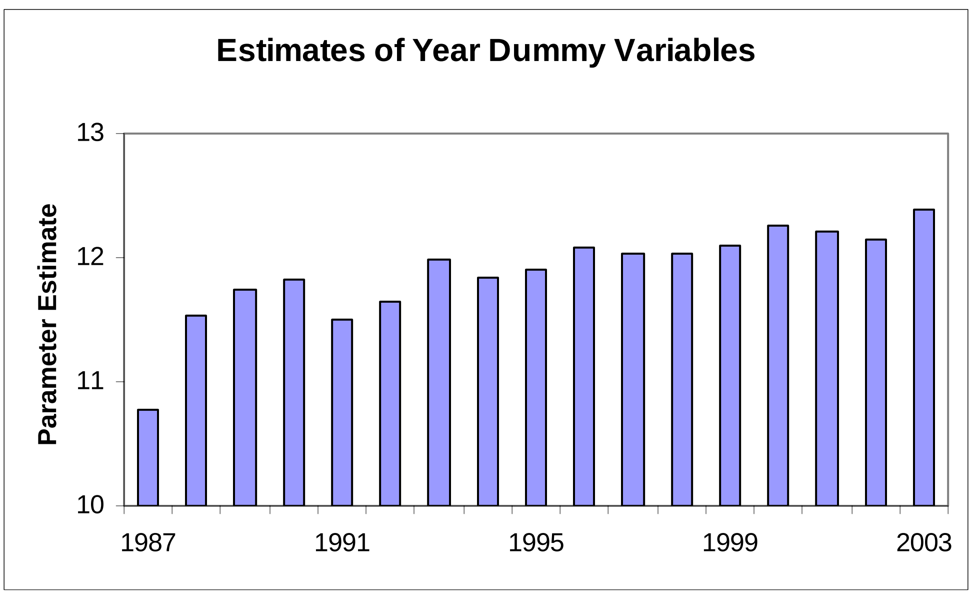

Figure 1: Estimated coefficients for the year dummy variables in Model 5, showing a discernible shift in the production frontier over time. All parameter estimates were statistically significant, and quite comparable among Models 3, 4, 5, and 6.

CONCLUSION

Not surprisingly, our results suggest that accounting for the panel structure of catch data in studies of fisheries efficiency can lead to quite different conclusions from those based on cross-sectional methods. Beyond this, we have found that representing the boat and time dimensions as two layers of a hierarchical model yields the best results among the models considered. Nevertheless, even when the models differed quite a bit by the measure of marginal likelihood, the results were broadly similar across most of the models. The most salient conclusions are that the elasticities of horsepower and tow hours are robustly positive across the models, and that there seems to have been a fairly steady outward shift in the technically efficient frontier over time. Estimation of the same family of models under translog production functions, which we have not reported on here due to space constraints, produced quite similar results.

Our estimates of mean efficiency are higher than those reported in most other studies. We note that these mean values obscure significant heterogeneity among individual units in the models, and particularly among the boat-year combinations of Models 5 and 6. We are currently investigating this heterogeneity, among other reasons for the purpose of testing whether there are statistically significant differences in efficiency among boats or years. One question along these lines is whether boats can be ranked by their efficiencies within and across years. We also hope to begin more general productivity studies to understand the components of productivity change in the Pacific hake fishery.

ACKNOWLEDGEMENTS

We thank Dave Colpo, William Daspit, and Brad Stenberg of the Pacific States Marine Fisheries Commission for help with data acquisition and interpretation; John Field, Tom Helser, Alec MacCall, Don Pearson, Steve Ralston, and Cindy Thomson for sharing their knowledge of the groundfish fishery; and participants at the NAAFE 2005 Forum for comments and suggestions.

REFERENCES

Aigner D, Lovell, K. and Schmidt, P., 1977. Formulation and Estimation of Stochastic Frontier Models. Journal of Econometrics 6, 21-37.

Alvarez, A., 2001. Some Issues in the Estimation of Technical Efficiency in a fishery. Efficiency paper series 2/2001, Department of Economics, University of Oviedo.

Chib, S., 1995. Marginal Likelihood From the Gibbs Output. Journal of the American Statistical Association 90, 1313-21.

Felthoven, R.G., Morrison Paul, C.J., 2004. Directions for productivity measurement in fisheries. Marine Policy 28, 161-169.

Fernandez, C., Ley, E., and Steel, M., 2002. Bayesian Modelling of Catch in a Northwest Atlantic Fishery. Journal of the Royal Statistical Society: Series C (Applied Statistics) 51, 257-280.

Helser, T., Methot, R., and Fleischer, G., 2004. Stock Assessment of Pacific Hake (Whiting) in U.S. and Canadian Waters in 2003. Pacific Fisheries Management Council.

Holloway, G. J., Tomberlin, D., and Irz, X., 2005. Hierarchical Analysis of Production Efficiency in a Coastal Trawl Fishery. In R. Scarpa and A. Alberini (eds.) Simulation Methods in Environmental and Natural Resource Economics, pp. 159-185, Kluwer Academic Publisher, Boston/Dordrecht/London.

Kumbhakar, S. C., and Lovell, C. A. K., 2000. Stochastic Frontier Analysis. Cambridge University Press, Cambridge.

Murillo-Zamorano, L.R. , 2004. Economic efficiency and frontier techniques. Journal of Economic Surveys 18, 33-78.

Walden, J. B., Kirkley, J. E., and Kitts, A. W., 2003. A Limited Economic Assessment of the Northeast Groundfish Fishery Buyout Program. Land Economics 79, 426-439.

Zellner A., 1996. An Introduction to Bayesian Inference in Econometrics. Wiley Classics Library. John Wiley and Sons, New York.