ESTIMATION OF A GENERALIZED FISHERY MODEL: A TWO-STAGE APPROACH

Junjie Zhang and Martin D. Smith*

Abstract—U.S. federal law calls for an end to overfishing, but measuring overfishing requires knowledge of bioeconomic parameters. Using microlevel economic data from the commercial fishery, this paper proposes a two-stage approach to estimate these parameters for a generalized fishery model. In the first stage, a fishery production function is consistently estimated by a within-period estimator treating the latent stock as a fixed effect. The estimated stock is then substituted into an equation of fish stock dynamics to estimate all other biological parameters. The bootstrap approach is used to correct the standard errors in the two-stage model. This method is applied to the reef-fish fishery in the northeastern Gulf of Mexico. The traditional method, which uses catch-per-unit-effort as a stock proxy, significantly overstates the optimal harvest level.

I. Introduction

THE reauthorization of the Magnuson-Stevens Fishery Conservation and Management Act in December 2006 "mandates that every fishery management plan contain an annual catch limit at a level such that overfishing does not occur in the fishery" (U.S. Department of Commerce, 2006). To determine what constitutes overfishing, fisheries managers require knowledge about biological processes and parameters. Because fish stocks are not directly observed and collecting biological data to assess stocks is costly, fishery managers often estimate biological parameters from economic data on catches and fishing inputs. With the increased demands of the reauthorized Magnuson-Stevens Act, managers are likely to rely even more heavily on fisheries data to assess fish stocks than they have in the past.

The standard approach for using fisheries data in stock assessment, which originated with Schaefer (1954), is to first specify an equation that relates output (catch) to fishing inputs (effort) and the latent fish abundance (stock), and then specify a state equation for biological dynamics in terms of the latent stock variable. Population parameters are thus recovered econometrically only from catch and effort data. However, the standard model introduces three empirical problems: (a) biological dynamics have natural variation for which models must account; (b) the production function has stochastic shocks, which makes the inferred stock noisy; and (c) the fishing production function that is used has an extremely restrictive form, namely Cobb-Douglas, with both exponents restricted to equal 1 (known commonly as the Schaefer harvest function). Ignoring any of these problems can lead to biased estimates of biological parameters and result in management targets that are set too high or too low.

Historically, inferring fish stocks from fishery data attracted economists and fisheries biologists because both professions are interested in determining how much of the natural resource can be extracted, and the problem involves an inseparable link between biological dynamics and an economic production function (Comitini & Huang, 1967; Pella & Tomlinson, 1969; Wilen, 1976). More recently, however, explorations of biological dynamics and economic production relationships have become separate endeavors. Biologists have focused on methods to address unobservable error terms in the stock dynamics and production functions simultaneously (de Valpine & Hastings, 2002; Schnute & Kronlund, 2002), while economists have focused on estimating production relationships conditioned on knowledge of stocks (Bjorndal & Conrad, 1987; Segerson & Squires, 1993).[1] As a result, no practical solution exists to address all three potential sources of bias simultaneously.

Modern bioeconomics recognizes that nonlinear systems are sensitive to their underlying parameters (Clark, 1990). Economists have emphasized the significance of biological parameter uncertainty for the fishing industry (Weitzman, 2002; Sethi et al., 2005). Even for a standard biological target, small differences in parameters can lead to large differences in the implied maximum sustainable yield (MSY). For an economic target, the optimal policy in dynamic fisheries models is a highly nonlinear function of the biological and economic parameters. In a nonlinear dynamic bioeconomic system, different parameters may imply different steady states. As a consequence, small parameter differences are amplified in the predicted or prescribed policy suggestion.

In this paper, we draw on panel econometric methods and propose a two-stage estimator that addresses all three problems with minimal computational burden. The key insight is that fishing vessels in each time period face the same stock. This allows us to difference away the unobserved stock and use a within-period estimator to produce consistent estimates of the production parameters in the first stage. We then estimate the stock from the production parameters, substitute it into the state equation that describes biological dynamics, and estimate biological parameters in a second step. We use

[1] Modern literature on capacity utilization has focused on refining techniques for estimating production frontiers and technical efficiency but has relegated inference about stock dynamics to fisheries scientists. See Kirkley, Paul, and Squires (2002) for a review of capacity utilization in fisheries. Estimated production relationships are not used to infer stocks. As such, there is a disconnect between the best description of production in economics and the description of production used in fishery science models of population dynamics.

the bootstrap method to derive correct standard errors for the two-stage approach.

We apply our model to fishing logbook data from the Gulf of Mexico reef-fish fishery and show that maximum sustainable yield for the fishery is biased upward by 77% if managers impose a simplified (Schaefer) harvest function. These results suggest that allowing separate estimates of biological parameters by biologists and production parameters by economists, as recent fisheries literature has done, may undermine sustainable fisheries management and ultimately lead to failures to achieve the stated goals of the Magnuson-Stevens Act.

In the next section, we discuss a generalized fishery model in detail. Section III introduces the two-stage estimation method and provides an illustration using a Monte Carlo experiment. Section IV is the empirical application to the Gulf of Mexico reef-fish fishery. Finally, we provide some brief conclusions in section V.

II.   Generalizing the Standard Model

The standard model used to infer logistic growth fish population dynamics using economic data from the fishery is a two-equation system (Schaefer, 1954). Let X denote the fish stock, H denote harvest or catch (which we use interchangeably at this point), and E denote fishing effort. The classical Gordon-Schaefer model is[2]

$$H_t = qE_tX_t, \quad \text{and} \tag{1}$$

$$X_{t+1} = X_t + rX_t\left(1 - \frac{X_t}{K}\right) - H_t, \tag{2}$$

where intrinsic growth r, carrying capacity K, and catchability q are unknown parameters.[3] If the stock, harvest, and effort information are all available, this structural model can be consistently estimated under regularity assumptions (Bjorndal & Conrad, 1987). However, detailed stock assessments, which are derived from surveys by fishery scientists, are not available for all species or collections of species. In this case, the biological stock is a latent variable to econometricians.

With unobserved stock, the empirical bioeconomic model is often estimated with proxies, such as the catch per skate used by Comitini & Huang (1967) and the average harvest-capital ratio used by Wilen (1976). Let y_t designate catch-per-unit-effort (CPUE), such that $y_t = H_t/E_t$. Then from equation (1), CPUE is proportional to the unobserved stock, that is, $X_t = y_t/q$. The standard approach recursively substitutes this proxy into equation (2) and appends an additive error term.

The result is an estimating equation for parameters r, q, and K (Pella & Tomlinson, 1969):

$$y_{t+1} = (1 + r)y_t - \frac{r}{qK}y_t^2 - qH_t + \varepsilon_t. \tag{3}$$

If the exogeneity condition holds such that $\mathbb{E}(\varepsilon_t|y_t, H_t) = 0$, this equation can be estimated by the least squares methods. We call this the CPUE estimator.

The potential drawbacks of using a production function that is homogeneous of degree 1 is widely acknowledged by economists but has largely been ignored in the fisheries science literature (Morey, 1986; Clark, 1990). Accounting for the law of diminishing returns, the Schaefer harvest function is a special case of the Cobb-Douglas production function,

$$H_t = qE_t^\alpha X_t^\gamma. \tag{4}$$

Models (2) and (4) can be estimated by the method similar to the CPUE estimator. The unobserved stock is now $X_t = (H_t/qE_t^\alpha)^{1/\gamma}$. With this proxy, the logistic growth model is reduced to the following estimating equation without stock:

$$\left(\frac{H_{t+1}}{qE_{t+1}^\alpha}\right)^{1/\gamma} = (1 + r)\left(\frac{H_t}{qE_t^\alpha}\right)^{1/\gamma} - \frac{r}{K}\left(\frac{H_t}{qE_t^\alpha}\right)^{2/\gamma} - H_t + \varepsilon_t. \tag{5}$$

With some distributional assumption on the error term, such as ε_t is independent and identically distributed (i.i.d.) normal with zero mean and variance σ_ε^2, this equation can be estimated through maximum likelihood (ML). We call this the CPUE-like estimator.

The CPUE-like estimator is straightforward but has two major problems. The first one is purely numerical; model (5) is highly nonlinear, so it is difficult for the estimator to converge (Tsoa, Schrank, & Roy, 1985). The estimation problem can be solved through Markov chain Monte Carlo (MCMC).[4] This tool is expedient for dealing with nonlinear models since it simulates instead of maximizing the likelihood function (Chernozhukov & Hong, 2003). Another concern is that the CPUE-like estimator incorporates the error in the stock dynamics but ignores the error in the production function (Polacheck, Hilborn, & Punt, 1993).[5] In particular, if the stock is measured with significant noise because of the unobservable shocks in the production function, the CPUE-like proxy will lead to biased estimators (Uhler, 1980).

Empirical bioeconomic studies involve three difficulties: the unobservable error terms in production functions and

---

[2] Gordon (1954) used this model structure in his seminal paper on common property resources. With the fishery production function proposed by Schaefer (1954), the fishery economics literature refers to it as a Gordon-Schaefer model.

[3] Intrinsic growth rate is the population growth rate between successive time periods. Carrying capacity is the maximum fish stock that can be supported by the ecosystem. It is also one of the steady states of the logistic growth function without harvest (another steady state is zero). In the Schaefer production function, catchability is the proportion of the fish stock removed by 1 unit of effort.

[4] This approach is adopted by Smith, Zhang, and Coleman (2007) to study a version of generalized Gordon-Schaefer model with aggregated time-series data. They show that this approach can recover structural parameters that classical estimation procedures fail to recover because they do not converge.

[5] Ecologists refer to the error term in the production function as observation error and that in the stock dynamics as process error. The standard fix for observation error is to assume that equation (2) holds without error and estimate equation (1) using maximum likelihood (Polacheck et al., 1993). Ecologists have only recently developed methods for addressing both process error and observation error (de Valpine & Hastings, 2002; Schnute & Kronlund, 2002), and these models are computationally intensive without adding generality to the production function in equation (1).

stock dynamics, the restrictive production function, and the latent stock. This paper proposes a new approach to deal with these problems simultaneously based on a generalized Gordon-Schaefer model. The generalized model includes multiple areas $j \in \{1 \cdots J\}$, multiple periods $t \in \{1 \cdots T\}$, multiple gears $g \in \{1 \cdots G\}$, and multiple vessels $i \in \{1 \cdots I\}$. The fishing vessels, gears deployed, and visited areas vary over time. The generalized Schaefer production function is

$$H_{ijgt} = q_{ijgt} E_{ijgt}^{\alpha_g} X_t^{\gamma} \exp(\epsilon_{ijgt}). \tag{6}$$

The stock is homogeneous across space for the sake of simplification. The restrictiveness of this assumption is relaxed by using heterogeneous catchability q_{ijgt} that is individual-, area-, gear- and time-specific. In model (6), the catch-effort elasticity α_g is gear-specific while the catch-stock elasticity γ is homogeneous. The stochastic term ϵ_{ijgt} reflects unobserved disturbances in the harvest process.

For the growth function, we distinguish harvest H and catch C. The aggregated harvest is less than or equal to total catch, that is, $\sum_{i,j,g} H_{ijgt} \leq C_t$, because fishers catch some nontargeted fish (known as bycatch).[6] The logistic growth function is given by

$$X_{t+1} = X_t + rX_t\left(1 - \frac{X_t}{K}\right) - C_t + \varepsilon_t. \tag{7}$$

The stochastic term ε_t is added to incorporate random ecological shocks to the biological system. The assumption of normality is added to conduct ML estimation but not necessary in least squares estimation.

Since the generalized Gordon-Schaefer model includes both error terms, methods based on just one type of error are problematic. We propose a two-stage estimation method as an alternative. The first stage is to estimate the stock from the production function. In the second stage, a regression is performed on the stock dynamics with estimated stock information. Because the estimated stock contains sampling error, the bootstrap approach is adopted to correct the standard errors. By this means, every parameter in equations (6) and (7) can be identified.

III.   The Two-Stage Estimation Method

A.   Estimating the Production Function

The first-stage estimation is concerned with the Cobb-Douglas production function (6). In this function, estimating an individual-, area-, gear- and time-specific catchability coefficient q_{ijgt} poses practical problems due to limited degrees of freedom, so we assume that

$$q_{ijgt} = \exp(\phi_g + a_j + \zeta_{ijgt}),$$

where ϕ_g is a gear-specific constant, a_j is an area-specific constant, and ζ_{ijgt} is a random error term to capture unobserved

[6] We define *bycatch* in section IV.

heterogeneity. The exponential function guarantees that the catchability is positive. From this point on, denote lower-case as the log transformation, for example, $x_t = \ln(X_t)$. The log-linearized production function is

$$h_{ijgt} = \phi_g + a_j + \alpha_g e_{ijgt} + \gamma x_t + \eta_{ijgt}, \tag{8}$$

where $\eta_{ijgt} = \zeta_{ijgt} + \epsilon_{ijgt}$. In this model, stock x_t is not observed by econometricians. Ignoring the latent stock will cause inconsistency because fishing effort e is correlated with the stock. If the stock randomly fluctuates in the neighborhood of the deterministic steady state such that $x_t \rightarrow x^*$ as $t \rightarrow \infty$, the stock can be written as $x_t = x^* + \varepsilon_t$ and $\mathbb{E}(\varepsilon_t | e_t, x_t) = 0$. In this case, the model can be estimated with a random effect model. However, assuming a steady state in each period is a very strong assumption, especially when there are constant shocks to the system, and it may take many periods to reach a steady state if biological or economic dynamics are slow.

Without the steady-state assumption, we can take advantage of the fishing data structure. In each period, different individuals with different characteristics share the same stock. If the period is chosen to be short enough, the stock could be treated as a constant (Bjorndal, 1989; Campbell, 1991). With this logic, a natural solution is to treat γx_t as a time effect and use the fixed-effect panel model. The constant stock can be canceled out through de-meaning among the fishing trips in the same period. Note that ϕ_g and a_j are gear- and area-specific constants. To avoid the dummy variable trap, we drop the dummy variables for gear type 1 (ϕ_1) and area 1 (a_1). Let $q = \exp(\phi_1 + a_1)$; then the time effect is defined as

$$c_t = \gamma x_t + \ln q. \tag{9}$$

The constant c_t can also be regarded as a stock index, since it is a linear function of log-transformed stock x_t. Two parameters γ and $\ln q$ scale the stock index such that changes in c_t signal changes in the underlying stock. To simplify notation, let z_{ijgt} be individual- and time-variant explanatory variables and β be a vector of parameters. Equation (8) is rewritten as

$$h_{ijgt} = z'_{ijgt}\beta + c_t + \eta_{ijgt}. \tag{10}$$

In this form, the generalized Schaefer harvest function is transformed to a panel data model with time-specific fixed effects. Under regularity conditions, model (10) can be consistently estimated by means of a within-period estimator. Let n be the number of cross-sectional units and T be the time periods. This estimator is consistent if either n or T, or both, tends to infinity (Hsiao, 2003). For fishery data, the panel is long in the sense of both large n and large T. In this case, the within estimator is $\sqrt{nT}$-consistent and asymptotically normal as $n, T \rightarrow \infty$.

We can identify all parameters in equation (8) except the time effect c_t, which is a function of stock x_t, the catch-stock elasticity γ, and two coefficients ϕ_1 and a_1. In the next

stage, we recover an estimate of stock and other remaining parameters using structural information contained in equation (7). Then we estimate the underlying logistic parameters that govern the stock dynamics.

B. Estimating the Growth Function

In this stage, we develop an estimation approach for the parameters in the logistic growth function (7), as well as γ, x, and q that are unidentified in the first stage. Note that the stock index c_t is a time-specific constant term in the production function (10). It can be estimated by

$$\hat{c}_t = \overline{h}_t - \overline{z}'_t\hat{\beta}, \tag{11}$$

where $\overline{h}_t = n_t^{-1} \sum_{i,j,g} h_{ijgt}$ and $\overline{z}_t = n_t^{-1} \sum_{i,j,g} z_{ijgt}$. Although $\hat{c}_t$ is unbiased, it is consistent only when $n_t \to \infty$. Otherwise if n_t is small, the estimated stock index contains significant noise. Aggregated appropriately in space and time, the fishery data satisfy this requirement. That is, there are a large number of fishing trips in the same period. The consistency of $\hat{c}_t$ is crucial for the second-stage estimation.

Although we are not able to recover stock in the previous regression, an index of stock abundance could be estimated through equation (11). When equations (9) and (11) are combined, the stock can be expressed as a function of $\hat{c}_t$ if parameters γ and q are known. Let $\hat{Y}_t = \exp(\hat{c}_t) = \exp(\overline{h}_t - \overline{z}'_t\hat{\beta})$; then the stock is estimated through

$$\hat{X}_t = \left[\frac{\exp(\overline{h}_t - \overline{z}'_t\hat{\beta})}{q}\right]^{1/\gamma} = \left(\frac{\hat{Y}_t}{q}\right)^{1/\gamma}. \tag{12}$$

Deriving $\hat{X}_t$ from equation (12), it is intuitive to run a regression on the logistic growth model (7), with X_t simply replaced by $\hat{X}_t$:

$$\left(\frac{\hat{Y}_{t+1}}{q}\right)^{1/\gamma} = (1 + r)\left(\frac{\hat{Y}_t}{q}\right)^{1/\gamma} - \frac{r}{K}\left(\frac{\hat{Y}_t}{q}\right)^{2/\gamma} - C_t + \varepsilon_t. \tag{13}$$

The above nonlinear model can be estimated through maximum likelihood estimation. However, there is a risk of multicollinearity because the aggregated variation of CPUE or CPUE-like proxy is very small (see section A.1 in appendix). To better identify the model, one approach is to restrict the catch-stock elasticity γ to 1. With this simplification and letting $s = r/(qK)$ and $\Delta \hat{Y}_{t+1} = \hat{Y}_{t+1} - \hat{Y}_t$, equation (13) can be linearized to

$$\Delta \hat{Y}_{t+1} = r\hat{Y}_t - s\hat{Y}_t^2 - qC_t + \varepsilon_t^*. \tag{14}$$

For this model, $\varepsilon_t^* = q\varepsilon_t$. This is a linear model with respect to $\hat{Y}_t$, $\hat{Y}_t^2$, and C_t, where r, s, and q can now be estimated through least square methods such as feasible generalized least square (FGLS). Let $\hat{\theta}^{LS}$ designate the parameter set $\{\hat{r}, \hat{s}, \hat{q}\}$ estimated by this method. If γ is restricted to 1, once q have been

identified, stock X_t can also be recovered through equation (12).

Up to this point, the two-stage estimator is similar to the methods using CPUE-like proxies, but it is superior.[7] First, this method uses the microlevel data and exploits the full information in the rich fishery data set. It also enables us to measure the variance of the estimated stock index. Second, some of the parameters are identified before estimating the logistic growth model, which reduces the computational burden in the nonlinear regression.

C. Consistency and Variance Estimation

The two-stage model described above is a special case of the multistep estimation approach. It is estimated sequentially because (a) the first-stage model is embedded in the second stage, (b) the joint distribution of this model is too complicated to derive so that full information maximum likelihood (FIML) is not feasible, and (c) the FIML introduces intensive computation in nonlinear models. The large sample properties of multistep estimators are discussed in detail by Newey and McFadden (1994).

In the previous section, the logistic growth model is estimated by replacing X_t with $\hat{X}_{\hat{t}}$. The consistency of $\hat{\theta}^{LS}$ hinges on a consistent estimate of $\hat{\beta}$, which holds under regularity conditions. For better understanding of this process, we derive the sampling error of the stock index. The estimated stock index $\hat{c}_t$ can be written as

$$\hat{c}_t = c_t + \overline{z}'_t(\beta - \hat{\beta}) + \overline{\eta}_t. \tag{15}$$

It is easy to show that the time effect is consistently estimated such that plim $\hat{c}_t = c_t$ because plim $\hat{\beta} = \beta$ and plim $\overline{\eta}_t = 0$. The consistency is guaranteed as long as the number of fishing trips in each period is large, that is, $n_t \to \infty$. We can also derive the variance of $\hat{c}_t$, which can be estimated by

$$\widehat{\text{var}(\hat{c}_t)} = \overline{z}'_t\left[\frac{1}{n_t}\widehat{\text{Avar}(\hat{\beta})}\right]\overline{z}_t + \frac{1}{n_t}\hat{\sigma}_\eta^2. \tag{16}$$

This equation shows that if the sample size (n_t) is small, the estimate of c_t contains significant noise. In this case, the small sample bias will be big. This phenomenon is similar to the case observed by Uhler (1980), which found that the optimal population size and harvest rate can be biased 40% to 50% depending on the model. We notice that this conclusion is drawn because aggregated production functions are used to infer stocks. The stocks contain measurement error, which causes bias in estimating biological parameters. When a microdata set with a very long time period and a large number of vessels is used, the bias will be diminishing. This is the case of our empirical data set in particular, and our approach using microdata will reduce bias in general.

---

[7] Note that with an aggregated data set and ignoring the error term in the production function, equation (13) is reduced to equation (5). Hence this is also a type of the CPUE-like estimator based on the microlevel data set.

TABLE 1.—ESTIMATION OF THE SIMULATED DATA SET

			First-Stage Estimate				Second-Stage Estimate		
			α_1	α_2	$\phi_2 - \phi_1$	$a_2 - a_1$	r	r/qK	q
T	n	True Value	0.3000	0.7000	−0.6000	0.3000	0.3000	0.1212	0.1237
50	50	Bias	0.0001	−0.0001	0.0007	−0.0001	0.0267	0.0102	0.0820
		RMSE	0.0031	0.0030	0.0160	0.0041	0.0394	0.0115	0.6561
	200	Bias	0.0000	0.0000	0.0002	−0.0001	0.0056	0.0029	−0.0067
		RMSE	0.0016	0.0016	0.0083	0.0019	0.0263	0.0039	0.2115
	1,000	Bias	0.0000	−0.0001	0.0002	0.0000	0.0018	0.0014	−0.0009
		RMSE	0.0007	0.0007	0.0035	0.0009	0.0353	0.0023	0.0539
200	50	Bias	0.0000	0.0001	−0.0003	−0.0001	0.0872	0.0352	0.0521
		RMSE	0.0015	0.0016	0.0085	0.0020	0.0897	0.0358	0.3974
	200	Bias	−0.0001	0.0000	−0.0003	0.0000	0.0253	0.0107	0.0026
		RMSE	0.0008	0.0007	0.0039	0.0010	0.0288	0.0110	0.0977
	1,000	Bias	0.0000	0.0000	0.0000	0.0000	0.0101	0.0052	0.0012
		RMSE	0.0004	0.0003	0.0019	0.0005	0.0186	0.0055	0.0224

Simulations are based on 5,000 replications. Catch-stock elasticity γ is restricted to 1, and $q = \exp(a_1 + \phi_1)$.

To illustrate how the small sample bias diminishes with sample size, we conduct a Monte Carlo experiment.[8] The artificial data set is generated from the data-generating processes (6) and (7). Fishing effort is randomly drawn from a uniform distribution $U[1, 100]$. The starting fish stock is half of the carrying capacity. The simulated data include 100 periods, 100 vessels, 2 gear types, and 2 areas. In each period, the vessels going fishing randomly choose one gear and one area. The true parameter values used in the generalized Gordon-Schaefer model are listed in the first row of table 1, as well as the two-stage estimates.

We report two statistics for the experiment. The first one, $\mathbb{E}(\hat{\theta} - \theta)$, measures the bias of the estimator. The second one is root-mean-squared error (RMSE) $\sqrt{\mathbb{E}(\hat{\theta} - \theta)^2}$ used to measure the efficiency. Each replication of the Monte Carlo experiment is repeated 5,000 times. The within-period estimator in the first-stage estimation is consistent without any assumption about the stock. The performance of the first-stage estimator is good even if the sample size is not huge. The second-stage estimate is subject to large bias when the sample size is small. The bias is especially sensitive to the number of cross-sectional units. This makes sense because the stock is estimated by vessels fishing in the same period. The efficiency also improves as the sample size grows.

The consistency of the two-stage approach can be justified with large samples under standard regularity conditions. However, the first-stage estimation will affect the asymptotic variance of the second stage. Standard errors in the second stage must be corrected to be used for statistical inference. One approach is to use the analytical corrected covariance matrix proposed by Murphy and Topel (1985). Another approach is to use bootstrapped standard errors. This paper adopts the panel bootstrap or block bootstrap technique. The paired bootstrap resamples over i but does not resample over t. The standard error derived in this way is asymptotically equivalent to the robust sandwich standard error (Kapetanios, 2008).

IV. The Reef-Fish Fishery in the Gulf of Mexico

The empirical study is based on the coastal reef-fish fishery in the northeastern Gulf of Mexico. Figure 1 depicts the study region and its thirteen fishing statistical zones. There are 62 commercially harvested reef-fish fishery species, including 11 grouper species and 10 snapper species. The reef-fish stock is assessed by the Stock Assessment Panel (1992–2003) and the Southeast Data Assessment and Review (SEDAR), pursuant to the Magnuson-Stevens Act.[9] These assessments show that most reef-fish stocks are overfished or undergoing overfishing. Especially red snapper, whose stock is comprehensively assessed, was severely overfished, and overfishing is still ongoing.

To address the overfishing problem, the Gulf of Mexico Fishery Management Council established a reef-fish fishery management plan (FMP) that was implemented in November 1984.[10] The goal of the FMP was "to manage the reef fish fishery of the United States waters of the Gulf of Mexico to attain the greatest overall benefits to the Nation with particular reference to food production and recreational opportunities on the basis of maximum sustainable yield (MSY) as modified by relevant economic, social, or ecological factors." Specifically, the FMP set collective reef-fish MSY to 51 million pounds per year.

The major data set used in this paper is the logbook data managed by the Southeast Fisheries Science Center, which consists of three relational databases. The vessel table contains vessel and general trip information, including a unique trip identifier, vessel ID, start and landing date of a trip, and fishing location. The gear table contains the amount and size of the gear reported for the trip, such as gear type and effort. The catch table includes the catches reported by the fishermen such as species and total weight.

There are 3,173 distinct vessels, 162,634 trips, and 404,908 trip days. The time span is from January 1, 1993, to October 31, 2004. We use only a subset of the data, specifically, the

---

[8] We also run two Monte Carlo experiments to show that the generalized model is indeed better than the restrictive Schaefer model. The result is available on request.

[9] http://www.sefsc.noaa.gov/sedar/.

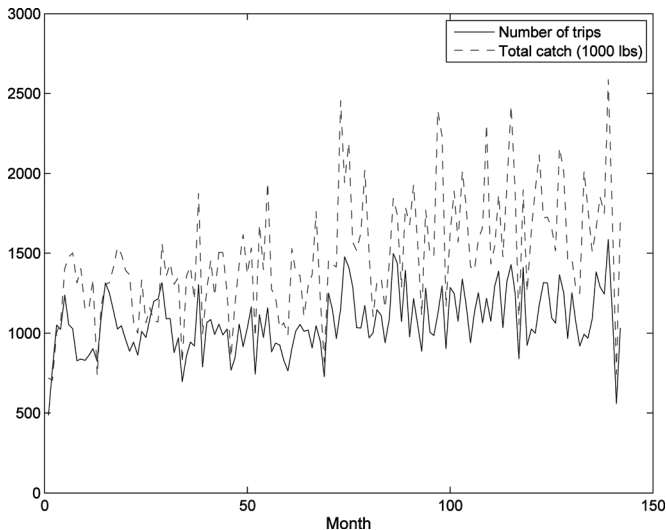[10] http://www.gulfcouncil.org/beta/gmfmcweb/FMPs/reef_fish_amend.htm.

FIGURE 1.—GULF OF MEXICO FISHERY STATISTICAL ZONES
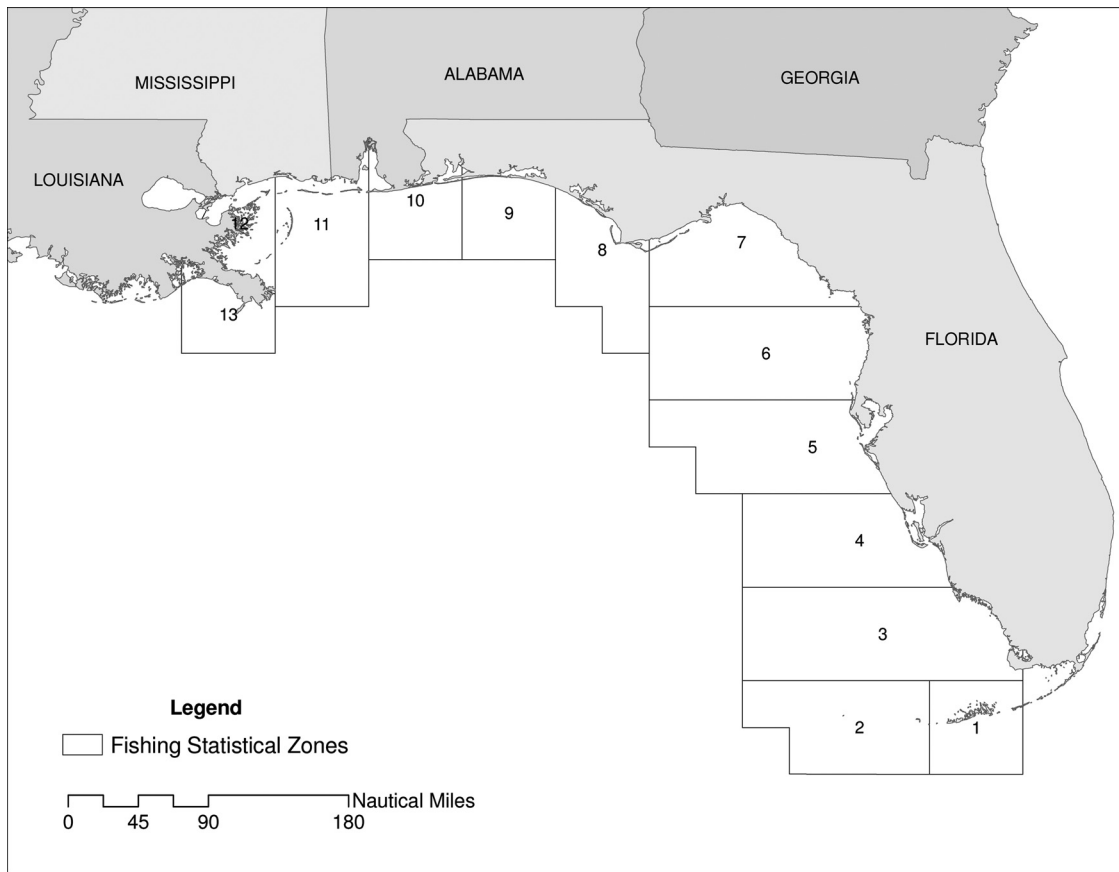
reef-fish trips from fishing statistical zones 1 to 13. To distinguish harvest from catch, we define a fishing trip as a reef-fish trip if the reef-fish catch accounts for at least 70 percent of the total catch. Otherwise, the reef-fish catch is regarded as bycatch. On average, reef-fish harvest accounts for 71% of total catch.

FIGURE 2.—AGGREGATED CATCH AND EFFORT IN EACH MONTH

The reef-fish fishery data are complicated because a single trip may contain multiple days, multiple areas, multiple species, and multiple gears. Besides defining reef-fish trips, we make the following assumptions. First, the effort is defined as number of crew times trip days. Second, if a trip involves multiple gears and areas, the catch and effort are assumed to be evenly distributed among each gear and area. Third, we lump all reef-fish species together. Fourth, the time step for stock dynamics is one month, and the stock is assumed to be the same to any fishermen during this period. Fifth, the stock is homogeneous across thirteen zones. In addition, observations with dubious information are dropped, such as crew number larger than ten or trip days longer than eight, based on our personal knowledge of the reef-fish fishery.

The final data set contains area-, month-, individual-, and gear-specific trip information. There are 142 months, 8 gear types, 13 areas, and 147,158 reef-fish fishing trips in total. The aggregated monthly trip numbers and total catch are depicted in figure 2. The dominant gear type is handline, accounting

TABLE 2.—SUMMARY STATISTICS OF THE REEF-FISH FISHERY DATA

Variable	Mean	sd	Minimum	Maximum
Harvest per trip (10^3 pounds)	0.9310	1.5610	0.001	111.2492
Effort per trip (crew × days)	26.9894	42.4568	1	640
CPUE per trip (harvest/effort)	0.1192	0.2851	0.0037	1.8675

FIGURE 3.—CPUE AND FISHING TRIPS FOR EACH GEAR AND EACH AREA
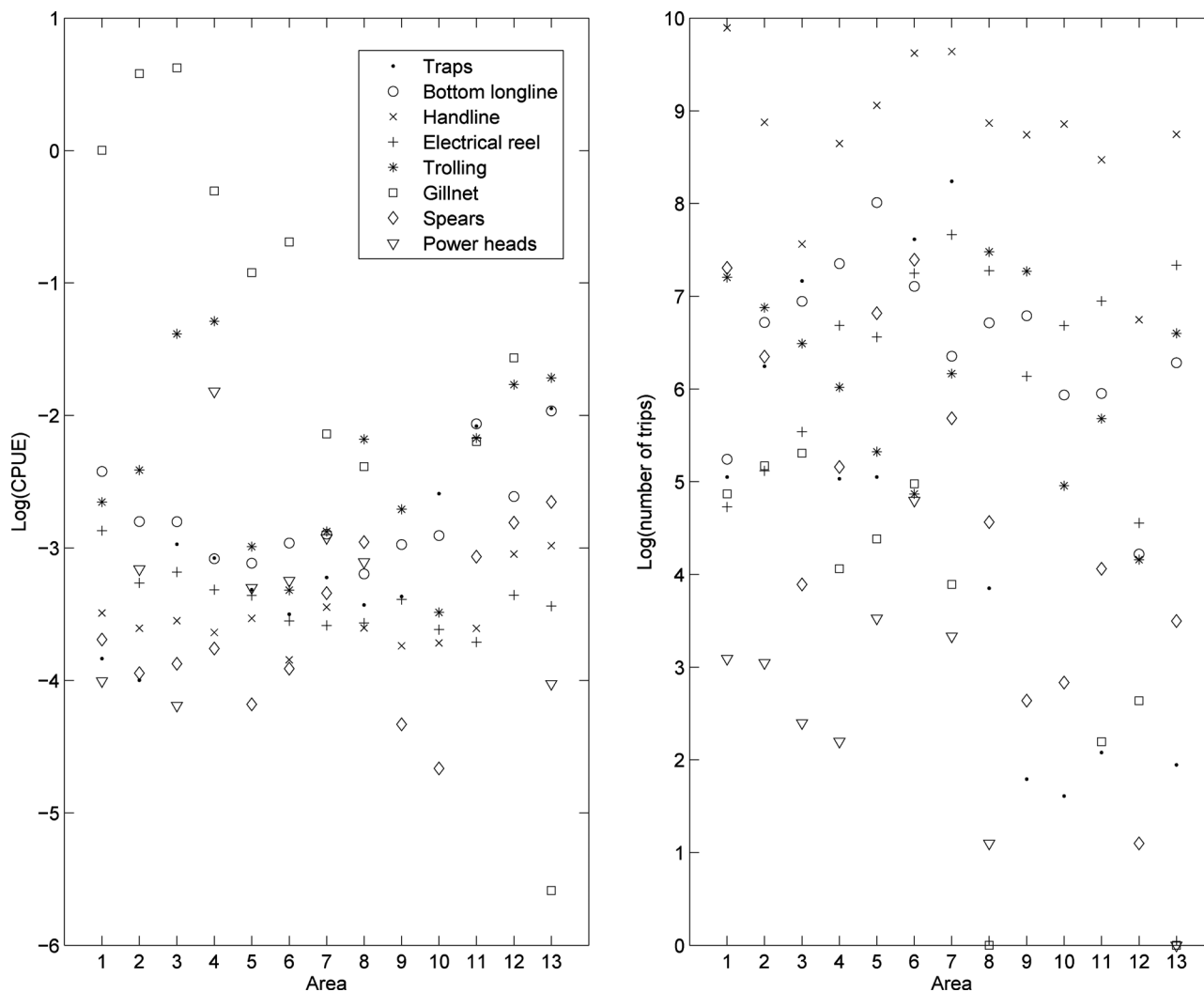
for 70 percent of total gears reported. Other gear types include traps, bottom longline, electrical reel/bandit rigs, trolling, gillnet, divers with spears, and divers with power heads. Summary statistics for the data set are compiled in table 2.

Figure 3 depicts CPUE and fishing trips for each gear in each area. The first panel shows that gillnet has the highest CPUE in areas 1 to 6 (the y axis is log-scaled). In fact, because gillnets are so effective, their use is closely monitored and regulated by the Fishery Management Council. For example, the gillnet fishery for Gulf king mackerel is subject to periodic closures. The second panel shows that handline accounts for the most fishing trips in all areas.

V.   Results and Discussion

Both the standard Gordon-Schaefer model and the generalized model are estimated in the empirical study. The bootstrapped standard errors are based on 1,000 block resamples over individuals but not over time. The results of the first-stage regression are reported in table 3. Every parameter estimate is significant at the 95% confidence level. Due to multicollinearity, the catchability for traps in area 1 is estimated as a reference case.

Comparing both models, we reject the specification of the Schaefer production function. The catch-effort elasticity is significantly less than 1, which means the reef-fish fishery is subject to the law of diminishing returns. Note that the conclusion that $\alpha_g < 1$ is true regardless of whether the catch-stock elasticity is restricted to 1, because γ is canceled out by the within-period de-meaning. The first-stage estimation is valid under standard regularity conditions, so the above conclusion holds without the second-stage regression and is independent of the model of population dynamics.

The first-stage estimates also show that there is significant heterogeneity over gears. As for the gear efficiency, the Schaefer model predicts that gillnet is by far the most efficient one, while trolling and bottom longline follow. However, the Schaefer model fails to account for the effect of effort on the gear efficiency. If the effort is the same, the generalized model compares the output of gears g and g' by

TABLE 3.—ESTIMATED PRODUCTION FUNCTION

Economic Parameters	Variable	Schaefer		Generalized Schaefer	
		Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Catch effort elasticity					
Traps	α_1			0.5026	0.0099
Bottom longline	α_2			0.3325	0.0093
Handline	α_3			0.6778	0.0021
Electrical reel/bandit rigs	α_4			0.6754	0.0062
Trolling	α_5			0.3956	0.0130
Gillnet	α_6			0.6717	0.0893
Divers with spears	α_7			0.7342	0.0128
Divers with power heads	α_8			0.6570	0.0574
Gear-specific constant					
Bottom longline	ϕ_2'	0.4476	0.0165	1.1287	0.0538
Handline	ϕ_3'	−0.2390	0.0119	−1.1189	0.0380
Electrical reel/bandit rigs	ϕ_4'	−0.3733	0.0160	−1.0032	0.0437
Trolling	ϕ_5'	0.7750	0.0182	−0.1709	0.0414
Gillnet	ϕ_6'	2.3897	0.0616	1.2880	0.1361
Divers with spears	ϕ_7'	−0.5760	0.0176	−1.6491	0.0478
Divers with power heads	ϕ_8'	0.0627	0.0623	−0.8842	0.1513
Area-specific constant					
Area 2	a_2'	0.3142	0.0149	0.6091	0.0124
Area 3	a_3'	0.5021	0.0175	0.7659	0.0157
Area 4	a_4'	−0.0538	0.0133	0.1764	0.0132
Area 5	a_5'	0.0635	0.0123	0.3905	0.0115
Area 6	a_6'	−0.0177	0.0105	0.3878	0.0099
Area 7	a_7'	0.3360	0.0105	0.6058	0.0093
Area 8	a_8'	0.0599	0.0124	0.4361	0.0122
Area 9	a_9'	−0.0467	0.0124	0.2278	0.0127
Area 10	a_{10}'	0.1720	0.0131	0.6330	0.0121
Area 11	a_{11}'	0.3528	0.0152	0.7603	0.0136
Area 12	a_{12}'	0.6695	0.0320	1.1246	0.0278
Area 13	a_{13}'	0.7775	0.0150	0.9708	0.0129

Total number of observations: 147,158. Time periods: 142. The catch-effort elasticity is restricted to 1 for the Schaefer harvest function. A prime denotes the normalized constant, such that $\phi_g' = \phi_g - \phi_1$ and $a_j' = a_j - a_1$. Bootstrapped standard errors are clustered on time.

$$\log h_g - \log h_{g'} = (\phi_g - \phi_{g'}) + (\alpha_g - \alpha_{g'}) \log E. \quad (17)$$

In equation (17), we ignore subscripts other than the gear since they are held fixed. Using the estimates in the first-stage regression, we plot the impact of gear choice on output in figure 4. Note that in this figure, trap is used as a reference gear. It also shows that gillnet is the most efficient gear, and its comparative efficiency increases over effort. This result is consistent with the gear-specific CPUE comparisons in the previous section. The gear with the highest catch-effort elasticity is divers with spears, and it is followed by handline, which is the dominant gear in the reef-fish fishery. The gear with the highest catchability is the gillnet, and the lowest one is divers with spears.

There is also observable spatial heterogeneity in the fishing production function. Using zone 1 as the reference area, the Schaefer model predicts that zone 13 has the highest catchability, while the generalized Schaefer model predicts that zone 12 is the highest one. Both models confirm that the zones lying to the south of Mississippi and Louisiana are easier to catch fish. Both models show that zone 4 has the lowest catchability.

After the first-stage regression, the stock index c_t can be estimated by equation (11). With the estimated stock information, we run a regression on the logistic growth model. To simplify the computation, the catch-stock elasticity γ is

FIGURE 4.—GEAR EFFICIENCY PREDICTED BY THE GENERALIZED MODEL
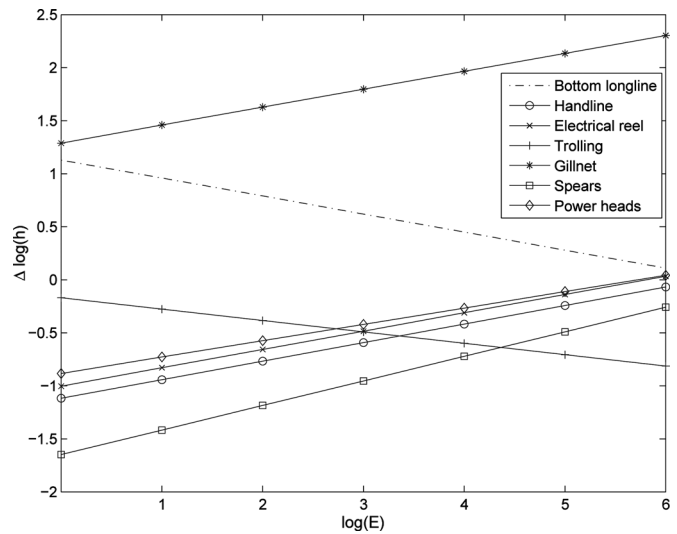
restricted to 1. The results of the second-stage estimation are reported in table 4. Since we have 142 periods and 147,158 observations in total, each period has 1,036 cross section units on average. The number of observations for each period is depicted in figure 2. A large number of active vessels fishing in the same period reduces the sampling errors of the stock index. A consistent estimate of stock enables us to obtain consistent estimates in the second stage.

In the second-stage regression, both Schaefer and generalized models yield significant estimates of biological parameters at the 95% confidence level. We use the bootstrap to estimate standard errors. Since we have rejected the Schaefer production function in the first stage, there is evidence that the Schaefer estimates in the second stage are biased. They underestimate the intrinsic growth rate and catchability while overestimating the lumped parameter.

Having the structural parameters identified, we provide some insights on the population dynamics. As a first step, we assume systems (6) and (7) are deterministic and use point estimates as the parameter values. That is, the parameter uncertainty is ignored. With this assumption, we analyze the implied carrying capacity and maximum sustainable yield.[11] The bottom two lines of table 4 report the result. The estimates of the Schaefer model are imprecisely estimated, while those of the generalized model are significant at the 95% confidence level.

Although the results are not significant, table 4 shows that the Schaefer regression overestimates the carrying capacity, and the implied maximum sustainable yield is too high. Moreover, the differences are not trivial. MSY, which is still the policy target for most fisheries, differs by as much as 77

[11] The maximum sustainable yield is the largest long-term catch that can be taken from a stock without impairing the biological system's renewability. In a Gordon-Schaefer model, the MSY is $rK/4$, and the corresponding stock producing MSY is $K/2$.

TABLE 4.—THE ESTIMATED LOGISTIC GROWTH FUNCTION

Name	Variable	Schaefer		Generalized Schaefer	
		Parameter Estimate	Standard Error	Parameter Estimate	Standard Error
Biological parameters					
Intrinsic growth rate	r	0.3442	0.0225	0.3847	0.0201
Lumped parameter	r/qK	9.0473	0.5126	2.5123	0.1824
Catchability	$q \times 10^{-5}$	0.0799	0.0322	0.5420	0.1307
Inferred parameters					
Carrying capacity	$K(\times 10^3 lbs)$	60,341.69	84,705.71	29,918.02	7,268.66
Maximum sustainable yield	$H^{MSY}(\times 10^3 lbs)$	5,064.72	6,069.76	2,860.81	630.80

Catch-stock elasticity γ is restricted to 1. Number of observations: 141. Bootstrapped standard errors.

percent. The estimate of Schaefer MSY is around 5.064 million pounds per month, which is higher than the collective MSY 4.25 million pounds per month (51 million pounds per year) originally set by the FMP. But for the estimates of the generalized model, the implied MSY is significantly smaller than the policy target (2.86 million pounds per month). The empirical results demonstrate that the restricted Schaefer production function causes serious problems when they are used to derive a steady-state management target in a deterministic bioeconomic system.

VI.   Conclusion and Future Research

This paper is an extension of previous research on estimating the Gordon-Schaefer type model. We propose a two-stage estimation method to tackle the latent stock problem with the presence of errors in the production function and stock dynamics. The approach is based on microlevel economic data. The within-period estimator in the first stage works well with standard regularity conditions. The estimated time effect is a good proxy to reflect the change in stock. The sampling error of the stock proxy will affect the second-stage covariance matrix, which is corrected by the bootstrap method. The advantages of the two-stage estimation method are to (a) harness the full information of fishing microdata, (b) provide a more general production function than the one traditionally used in estimating stock dynamics, and (c) implement in a computationally simple manner.

This approach is applied to the reef-fish fishery in the Gulf of Mexico. The empirical study shows that the restrictive Schaefer production function produces biased estimates, which imply large differences in the steady state of the dynamic system. The empirical study shows that the Schaefer regression significantly overestimates the carrying capacity, as well as the maximum sustainable yield. Fishery management based on these different estimates would differ dramatically. Taken to its logical extreme, pursuit of an MSY target that is set too high will eventually collapse the fishery altogether.

While the proposed two-stage approach makes empirical study of a standard fishing model computationally feasible and consistently estimated, it has some limitations. First, the catch-stock elasticity is restricted to 1 to reduce nonlinearity. Although we can readily relax this assumption and adopt the

maximum likelihood estimation, this generalization significantly increases the computational intensity and increases multicollinearity.

Our method provides a straightforward and computationally feasible approach to estimate stock dynamics without the substantial cost of collecting fishery-independent data. This will help to fulfill one promise of the reauthorized Magnuson-Stevens Act. However, the act also calls for the use of ecosystem approaches in fisheries management. This requirement will inevitably introduce the need to analyze spatially explicit stock dynamics. Theoretical work on optimal spatial management of fisheries is beginning to emerge (Sanchirico & Wilen, 2005; Costello & Polasky, 2008). Thus, extending our approach to the spatial domain is important for future research.

An important caveat for our empirical application is that we omit recreational fishery data. In many U.S. fisheries, recreational catches account for a significant portion of total landings (Coleman et al., 2004). We acknowledge that omitting recreational fishing could lead to bias, but we are unable to correct for this bias with the available data. Logbook-style data are not available for our study area with spatial resolution in the different fishing zones. We suggest that state and federal government establish a better recreational data collection system that is comparable to that used for commercial fisheries. A coalition of recreational fishing interests has proposed such a data collection system for Gulf of Mexico reef fish.[12]

REFERENCES

Bjorndal, T., "Production in a Schooling Fishery: The Case of the North-Sea Herring Fishery," *Land Economics* 65:1 (1989), 49–56.
Bjorndal, T., and J. M. Conrad, "The Dynamics of an Open Access Fishery," *Canadian Journal of Economics* 20:1 (1987), 74–85.
Campbell, H. F., "Estimating the Elasticity of Substitution between Restricted and Unrestricted Inputs in a Regulated Fishery: A Probit Approach," *Journal of Environmental Economics and Management* 20:3 (1991), 262–274.
Chernozhukov, V., and H. Hong, "An MCMC Approach to Classical Estimation," *Journal of Econometrics* 115:2 (2003), 293–346.
Clark, C. W., *Mathematical Bioeconomics: The Optimal Management of Renewable Resources*, 2nd ed. (New York: Wiley, 1990).

[12] Recreational Fishery Management Proposal for Gulf of Mexico Reef Fish: http://www.gulfcouncil.org.

Coleman, F. C., W. F. Figueira, J. S. Ueland, and L. B. Crowder, "The Impact of United States Recreational Fisheries on Marine Fish Populations," *Science* 305:5692 (2004), 1958–1960.

Comitini, S., and D. S. Huang, "Study of Production and Factor Shares in Halibut-Fishing-Industry," *Journal of Political Economy* 75:4 (1967), 366–372.

Costello, C., and S. Polasky, "Optimal Harvesting of Stochastic Spatial Resources," *Journal of Environmental Economics and Management* 56:1 (2008), 1–18.

Davidson, Russell, and J. G. MacKinnon, *Estimation and Inference in Econometrics* (New York: Oxford University Press, 1993).

Gordon, H. S., "The Economic Theory of a Common-Property Resource: The Fishery," *Journal of Political Economy* 62:2 (1954), 124–142.

Hsiao, C., *Analysis of Panel Data*, 2nd ed. (Cambridge: Cambridge University Press, 2003).

Kapetanios, G., "A Bootstrap Procedure for Panel Data Sets with Many Cross-Sectional Units," *Econometrics Journal* 11:2 (2008), 377–395.

Kirkley, J., C. J. M. Paul, and D. Squires, "Capacity and Capacity Utilization in Common-Pool Resource Industries: Definition, Measurement, and a Comparison of Approaches," *Environmental and Resource Economics* 22:1–2 (2002), 71–97.

Morey, E. R., "A Generalized Harvest Function for Fishing: Allocating Effort among Common Property Cod Stocks (a Generalized Harvest Function)," *Journal of Environmental Economics and Management* 13:1 (1986), 30–49.

Murphy, K. M., and R. H. Topel, "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics* 3:4 (1985), 370–379.

Newey, W. K., and D. L. McFadden, "Large Sample Estimation and Hypothesis Testing" (pp. 2111–2245), in R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics* (New York: Elsevier Science, 1994).

Pella, J. J., and P. K. Tomlinson, "A Generalized Stock Production Model," *Bulletin of the Inter-American Tropical Tuna Commission* 13:3 (1969), 421–454.

Polacheck, T., R. Hilborn, and A. E. Punt, "Fitting Surplus Production Models: Comparing Methods and Measuring Uncertainty," *Canadian Journal of Fisheries and Aquatic Sciences* 50:12 (1993), 2597–2607.

Sanchirico, J., and J. Wilen, "Managing Renewable Resource Use with Market-Based Instruments: Matching Policy Scope to Ecosystem Scale," *Journal of Environmental Economics and Management* 50:1 (2005), 23–46.

Schaefer, M. B., "Some Aspects of the Dynamics of Populations Important to the Management of Commercial Marine Fisheries," *Bulletin of the Inter-American Tropical Tuna Commission* 1:2 (1954), 26–56.

Schnute, J. T., and A. R. Kronlund, "Estimating Salmon Stock-Recruitment Relationships from Catch and Escapement Data," *Canadian Journal of Fisheries and Aquatic Sciences* 59:3 (2002), 433–449.

Segerson, K., and D. Squires, "Capacity Utilization under Regulatory Constraints," this REVIEW 75:1 (1993), 76–85.

Sethi, G., C. Costello, A. Fisher, M. Hanemann, and L. Karp, "Fishery Management under Multiple Uncertainty," *Journal of Environmental Economics and Management* 50:2 (2005), 300–318.

Smith, M. D., J. Zhang, and F. C. Coleman, "Structural Modeling of Marine Reserves with Bayesian Estimation," *Marine Resource Economics* 22:2 (2007), 121–136.

Tsoa, E., W. E. Schrank, and N. Roy, "Generalizing Fisheries Models: An Extension of the Schaefer Analysis," *Canadian Journal of Fisheries and Aquatic Sciences* 42:1 (1985), 44–50.

Uhler, R. S., "Least-Squares Regression Estimates of the Schaefer Production-Model: Some Monte-Carlo Simulation Results," *Canadian Journal of Fisheries and Aquatic Sciences* 37:8 (1980), 1284–1294.

U.S. Department of Commerce, "Congress Approves Magnuson-Stevens Fishery Conservation and Management Act," press release December 9, 2006. http://commerce.senate.gov.

Valpine, P. de, and A. Hastings, "Fitting Population Models Incorporating Process Noise and Observation Error," *Ecological Monographs* 72 (2002), 57–76.

Weitzman, M. L., "Landing Fees vs Harvest Quotas with Uncertain Fish Stocks," *Journal of Environmental Economics and Management* 43:2 (2002), 325–338.

Wilen, J. E., "Common Property Resources and the Dynamics of Overexploitation: The Case of the North Pacific Fur Seal," University of British Columbia, Natural Resource Economics working paper no. 3 (1976).

APPENDIX

Maximum Likelihood Estimation of the Second-Stage Model

If ε_t is assumed to be i.i.d. normally distributed with mean zero and variance σ_ε^2, equation (13) can be estimated through maximum likelihood estimation. In addition, we assume that the error terms in the production function and the logistic growth function are independent. If the distribution is correctly specified, the ML estimator has many desirable properties. Besides consistency and asymptotic normality, it is invariant to reparameterization. Because we are interested in the original parameter instead of the reparameterized one, this property is important. Let θ^{MLE} be a vector of reparameterized parameters: $\theta_1 = q^{-1/\gamma}$, $\theta_2 = (1+r)q^{-1/\gamma}$, $\theta_3 = rq^{-2/\gamma}K^{-1}$, and $\theta_4 = \gamma^{-1}$, then the log-likelihood function is

$$\ln \mathcal{L}(\theta, \sigma^2) = -\frac{T-1}{2}\ln(2\pi\sigma^2)$$
$$-\frac{1}{2\sigma^2}\sum_t \left[\theta_1 \hat{Y}_{t+1}^{\theta_4} - \left(\theta_2 \hat{Y}_t^{\theta_4} - \theta_3 \hat{Y}_t^{2\theta_4} - C_t\right)\right]^2. \quad (A.1)$$

The nonlinear model (13) is straightforward but may have an identification problem in estimating the Gordon-Schaefer model. An important feature of fishery data is that the aggregated variation of CPUE or CPUE-like proxy is very small, which causes multicollinearity. Let $R_t = \theta_1 \hat{Y}_{t+1}^{\theta_4} - (\theta_2 \hat{Y}_t^{\theta_4} - \theta_3 \hat{Y}_t^{2\theta_4} - C_t)$. Then the first-order condition of the log likelihood function (A.1) with respect to θ is

$$\frac{\partial \ln \mathcal{L}}{\partial \theta_1} = -\frac{1}{2\sigma^2}\sum_t 2R_t \hat{Y}_{t+1}^{\theta_4} = 0,$$

$$\frac{\partial \ln \mathcal{L}}{\partial \theta_2} = \frac{1}{2\sigma^2}\sum_t 2R_t \hat{Y}_t^{\theta_4} = 0,$$

$$\frac{\partial \ln \mathcal{L}}{\partial \theta_3} = -\frac{1}{2\sigma^2}\sum_t 2R_t \hat{Y}_t^{2\theta_4} = 0, \quad \text{and}$$

$$\frac{\partial \ln \mathcal{L}}{\partial \theta_4} = -\frac{1}{2\sigma^2}\sum_t 2R_t \left[\theta_1 \hat{Y}_{t+1}^{\theta_4} \ln \hat{Y}_{t+1} - \theta_2 \hat{Y}_t^{\theta_4} \ln \hat{Y}_t + 2\theta_3 \hat{Y}_t^{2\theta_4} \ln \hat{Y}_t\right] = 0.$$

Note that if the variation of $\hat{Y}_t$ is very small over time, $\partial \ln \mathcal{L}/\partial \theta_1$ is close to $\partial \ln \mathcal{L}/\partial \theta_2$ in absolute value. In addition, $\partial \ln \mathcal{L}/\partial \theta_4$ is very close to a linear combination of the other derivatives. Therefore, the first-order condition may suffer from singularity and cause a multicollinearity problem (Davidson & MacKinnon, 1993).