ESTIMATING A GENERALIZED GORDON-SCHAEFER MODEL WITH HETEROGENEOUS FISHING DATA

Junjie Zhang  
Nicholas School of the Environment & Earth Sciences, Duke University  
junjie.zhang@duke.edu  

Martin D. Smith  
Nicholas School of the Environment & Earth Sciences, Duke University  

Abstract  
This paper proposes a two-stage method to estimate a generalized Gordon-Schaefer model, using heterogeneous fishing data in the logbook. A Cobb-Douglas production function is consistently estimated in the first stage by means of the ‘within period estimator’. A stock index is constructed and used to estimate the logistic growth model in the second stage. Realizing the estimated stock index contains measurement error, the simulation-extrapolation method is used to correct bias. From this, all parameters as well as the latent stock can be identified. This method is also applied to the reef-fish fishery in the northeastern Gulf of Mexico. The traditional method, which uses catch-per-unit-effort (CPUE) as a proxy and is ignoring measurement error, significantly overstates the carrying capacity and the maximum sustainable yield.

Keywords: Gordon-Schaefer model, latent variable, panel data, measurement error

INTRODUCTION

Optimal management of fisheries requires knowledge about biological processes and parameters. The Gordon-Schaefer model is widely adopted to analyze fishery management and policy. Its popularity is partly due to its simplicity, which makes the optimal control problem tractable. However, the Gordon-Schaefer model is restrictive for three reasons. First, the Schaefer production function ignores the law of diminishing returns in a fishery because it assumes harvest is homogeneous of degree one with respect to effort and stock (Morey, 1986; Coppola and Pascoe, 1998). Second, the production function cannot be estimated directly unless the stock information is available (Bjorndal and Conrad, 1987). However, detailed stock assessments, which are derived from surveys by fishery scientists, are not always available for all species. Third, there are unobserved disturbances to the production and biological processes, and the specification of error terms may result in different estimators (Polacheck et al., 1993). The first problem could be solved by introducing a nonlinear catch-effort relationship, but the other two problems deserve more attention.

The unavailable stock information makes some empirical studies infeasible, so appropriate assumptions or simplifications must be made. With aggregated time series data, the catch-per-unit-effort (CPUE) is often used as a proxy for stock, such as the ‘catch per skate’ used by Comitini and Huang (1967) and the averaged harvest/capital ratio used by Wilen (1976). A similar method is to adopt an unrestricted Cobb-Douglas function, and express stock as a nonlinear function of catch and effort. Using this CPUE-like proxy, the generalized Gordon-Schaefer model is transformed to a reduced form through recursive substitution, and it could be estimated by classical or Bayesian methods (Tsoa et al., 1985; Smith et al., 2005). With cross-sectional data, stock can be assumed to be common to each fishing unit and be treated as a constant (Bjorndal, 1989). Acknowledging that fish stocks may evolve during a fishing season, Campbell (1991) uses a monthly dummy variable
to accommodate stock fluctuation. With longitudinal data, Smith et al. (2006) adopt a log-linear biological model. This approach greatly simplifies estimation and allows a large number of control variables to be included. Although it is convenient for backward-looking policy evaluation, this approach may oversimplify the stock dynamics and make long-run policy analysis inappropriate.

In the joint estimation of production and growth models, using CPUE-like proxy is the most popular method. However, this method is problematic. One problem is purely numerical: the reduced form model is highly nonlinear, so it is difficult for the estimator to converge (Tsoa et al., 1985). A more pressing concern is the ‘errors-in-variables’ problem. Because there are unobserved disturbances in the production function, the CPUE-like proxy contains measurement error, which will bias the estimator (Uhler, 1980; De Valpine and Hastings, 2002). We focus on classical estimation approaches in this paper, but one could also address both numerical and the errors-in-variables problem with Bayesian methods (Schnute and Kronlund, 2002; De Valpine and Hilborn, 2005).

The purpose of this paper is to estimate a generalized Gordon-Schaefer model without stock information. The two-stage estimation method proposed is also a generalization of previous studies, taking into account the specification of disturbance terms and measurement error. This method is based on individual logbooks, which record individual vessel, gear, fishing area, number of crew, and time-specific fishing trips. With longitudinal data, we are able to recover parameters that cannot be identified by cross-sectional or time series data (Wooldridge, 2002). A Monte Carlo experiment is conducted to test this method, but it is not reported due to the space constraint. We provide an empirical application to the Gulf of Mexico reef-fish fishery.

In the next section we discuss the generalized Gordon-Schaefer model in detail. The third section introduces the two-stage estimation method. The following section is the empirical application to the Gulf of Mexico reef-fish fishery. Finally, we provide some brief conclusions.

THE MODEL

The generalized Gordon-Schaefer model developed in this paper includes one area, multiple period \((t)\), multiple gears \((k)\) and multiple vessels \((i)\). The number of vessels and gear types on the sea during the same period vary over time. Let \(I(t)\) be a set including individuals fishing at time \(t\), and \(I(t) \subseteq \{1 \cdots I\}\), where \(I\) is the maximum number of vessels going fishing at the same time. Let \(K(t)\) be a set including gear types used at time \(t\), and \(K(t) \subseteq \{1 \cdots K\}\), where \(K\) is the maximum type of gears. The time period is from 1 to \(T\). Denote \(H\) as target catch, \(C\) as total catch, \(E\) as fishing effort, and \(X\) as stock. Because of the bycatch problem, the aggregated target catch is less than or equal to total catch, that is, \(\sum_{i} \sum_{k} H_{ikt} \leq C_t\). The generalized Gordon-Schaefer model is given by (Gordon, 1954; Schaefer, 1954; Tsoa et al., 1985; Clark, 1990)

\[
H_{ikt} = q_{ikt} E_{ikt}^{\alpha_{1k}} X_t^{\alpha_{2}} \exp(\epsilon_{ikt}), \quad \text{and} \quad (1)
\]

\[
X_{t+1} = X_t + rX_t \left(1 - \frac{X_t}{S}\right) - C_t + \varepsilon_t, \quad i \in I(t), \ k \in K(t), \ t = 1 \cdots T. \quad (2)
\]

Equation (1) is a Cobb-Douglas production function, where \(q_{ikt}\) is the individual-, gear- and time-specific catchability coefficient. For the estimation to be practicable, some simplifications will be made on \(q_{ikt}\) later. The catch-effort elasticity is \(\alpha_{1k}\), which is gear-specific, while catch-stock elasticity \(\alpha_{2}\) is homogeneous. To reduce nonlinearity, \(\alpha_{1}\) or \(\alpha_{2}\) could be restricted to one. Note that \(\alpha_{2} = 1\) is a convenient assumption, which is adopted in the empirical study. The error term
\( \epsilon_{ikt} \) reflects unobserved disturbances in the harvest process, which are assumed to be iid with mean zero and variance \( \sigma^2_\epsilon \). Equation (2) is a logistic growth model, where \( r \) is the intrinsic growth rate and \( S \) is the carrying capacity. The disturbance term \( \epsilon_t \) is added to incorporate random ecological shocks to the biological system. It is assumed to be iid normal with mean zero and variance \( \sigma^2_\epsilon \).

The normality assumption is added to conduct the maximum likelihood estimation. The setup of error terms is comparable to the ‘observation-error estimators’ and ‘process-error estimators’ defined by Polacheck et al. (1993). The first one refers to the error in the relationship between catch, effort and stock, and the second one refers to the error in the change in population size.

Since the generalized Gordon-Schaefer model includes both ‘observation error’ and ‘process error’, 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 a stock abundance index. In the second stage, a regression is performed on the stock dynamics with estimated stock information. Because the estimated stock contains errors, some correction must be made to reduce the bias due to the measurement error. By this means, every parameter in equation (1) and (2) can be identified.

**THE TWO-STAGE ESTIMATION METHOD**

**The First Stage**

The first stage estimation is concerned with the Cobb-Douglas function. Denote tilde as the log transformation, for example, \( \tilde{X}_t \equiv \log(X_t) \). In this paper, the log-transformed catchability coefficient \( \tilde{q}_{ikt} \) is assumed to be an iid random variable with mean \( \tilde{q}_k \). It can be written as \( \tilde{q}_{ikt} = \tilde{q}_k + e_{ikt} \), where \( e_{ikt} \) is an iid random variable with mean zero and variance \( \sigma^2_e \). This setup implies that individual heterogeneity is captured by \( e_{ikt} \). In this form, \( e_{ikt} \) can be combined with \( \epsilon_{ikt} \), so let \( \eta_{ikt} \equiv e_{ikt} + \epsilon_{ikt} \), which is also iid with mean zero and variance \( \sigma^2_\eta \).

Taking the log transformation, production function (1) is linearized as

\[
\tilde{H}_{ikt} = \tilde{q}_k + \alpha_1 k \tilde{E}_{ikt} + \alpha_2 \tilde{X}_t + \eta_{ikt}, \quad i \in I(t), \; k \in K(t), \; t = 1 \cdots T. \tag{3}
\]

In this model, stock \( \tilde{X}_t \) is a latent variable, since it cannot be observed directly. One natural treatment is to include time-specific dummy variables and run a least squares regression, then \( \alpha_2 \tilde{X}_t \) is estimated as a set of constants. However, the least squares dummy variable (LSDV) regression method is not practically feasible if the time period is very long, because there are too many regressors (Wooldridge, 2002). Another problem is that \( \alpha_2 \) and \( \tilde{X}_t \) cannot be separately identified by equation (3) alone if \( \tilde{X}_t \) is treated as a constant. In fact, even \( \alpha_2 \tilde{X}_t \) cannot be identified with one gear-specific catchability coefficient due to the ‘dummy variable trap’ (will be explained below).

Inspired by panel data estimation methods, we could take advantage of the fishing data structure. In each period, different individuals with different gears fish in the same area, which means they share the same stock. If the period is chosen to be short enough, the stock could be treated as a constant. This assumption is consistent with the observation by Bjorndal (1989) and Campbell (1991). The constant stock can be canceled out through demeaning among the fishing trips in the same period and same area. Let \( n_t \) be the total number of fishing trips at time \( t \), \( m_{kt} \) be the number of gear type \( k \) deployed at time \( t \), where \( n_t = \sum_k m_{kt} \). As discussed before, which individuals are fishing and which gears are being deployed both vary over time. To simplify notation, \( i \) and \( k \) in the summation below indicate \( i \in I(t) \) and \( k \in K(t) \) respectively. Averaging the log transformed
production function (3) over gears and individuals in each period, we derive

\[ \frac{1}{n_t} \sum_k \sum_i \tilde{H}_{ikt} = \frac{1}{n_t} \sum_k m_{kt} \tilde{q}_k + \frac{1}{n_t} \sum_k \left( \alpha_k + \tilde{E}_{ikt} \right) + \alpha_2 \tilde{X}_t + \frac{1}{n_t} \sum_i \eta_{ikt}. \] (4)

Since \( n_t = \sum_k m_{kt} \), the number of gear type \( j \) deployed at time \( t \) can be determined by \( n_t \) and the total number of other gear types deployed, that is, \( m_{jt} = n_t - \sum_{k \neq j} m_{kt} \). Without loss of generality, let \( j = 1 \), then the first item on the RHS of equation (4) is now

\[ \frac{1}{n_t} \sum_k m_{kt} \tilde{q}_k = \tilde{q}_1 + \frac{1}{n_t} \sum_k m_{kt} (\tilde{q}_k - \tilde{q}_1). \] (5)

By subtracting equation (4) from (3), periodic demeaning cancels out the stock \( X_t \), the catch-stock elasticity \( \alpha_2 \) and the catchability coefficient \( q_1 \). In this way, the pooled ordinary least square (OLS) has the remaining parameters consistently estimated on the demeaned data. To simplify notation, let \( \beta \) be a parameter vector \([\tilde{q}_1, \ldots, \tilde{q}_K \tilde{q}_1, \alpha_{11}, \ldots, \alpha_{1K}]^\prime \). Let \( \mathbf{Z}_{ikt} \) be a collection of data including gear dummy variables and effort information \([0, \ldots, 1, \ldots, 0, 0, \ldots, \tilde{E}_{ikt}, \ldots, 0]^\prime \).

The first \( K - 1 \) elements are gear-specific dummy variables, and the last \( K \) elements are log-transformed effort. If an individual vessel \( i \) chooses gear \( k \) (\( k \geq 2 \)) at time \( t \), then the \((k - 1)\)th element is 1 and the \((K - 1 + k)\)th element is \( \tilde{E}_{ikt} \). Equation (3) is rewritten as

\[ \tilde{H}_{ikt} = \tilde{q}_1 + \mathbf{Z}_{ikt} \beta + \alpha_2 \tilde{X}_t + \eta_{ikt}, \quad i \in \mathcal{I}(t), \quad k \in \mathcal{K}(t), \quad t = 1 \cdots T. \] (6)

Also define \( \overline{H}_t = n_t^{-1} \sum_k \sum_i \tilde{H}_{ikt}, \overline{\mathbf{Z}}_t = n_t^{-1} \sum_k \sum_i \mathbf{Z}_{ikt} \) and \( \overline{\eta}_t = n_t^{-1} \sum_k \sum_i \eta_{ikt} \). Then the averaged equation (4) can now be simplified as

\[ \overline{H}_t = \tilde{q}_1 + \overline{\mathbf{Z}}_t \beta + \alpha_2 \overline{X}_t + \overline{\eta}_t, \quad t = 1 \cdots T. \] (7)

Subtracting equation (7) from equation (3), the stock \( X_t \), catch-stock elasticity \( \alpha_2 \) and catchability coefficient \( \tilde{q}_1 \) cancel. Defining \( \tilde{H}_{ikt} = \tilde{H}_{ikt} - \overline{H}_t \) and \( \tilde{\eta}_{ikt} = \eta_{ikt} - \overline{\eta}_t \). Let \( \tilde{\mathbf{Z}}_{ikt} = \mathbf{Z}_{ikt} - \overline{\mathbf{Z}}_t \) be a collection of demeaned gear and effort information, then the periodic demeaned equation is now

\[ \tilde{H}_{ikt} = \tilde{\mathbf{Z}}_{ikt} \beta + \tilde{\eta}_{ikt}, \quad i \in \mathcal{I}(t), \quad k \in \mathcal{K}(t), \quad t = 1 \cdots T. \] (8)

For this model, the parameter set \( \beta \) can be consistently estimated through pooled OLS since \( \mathbb{E}(\eta | \tilde{Z}) = 0 \). Because the estimator \( \tilde{\beta} \) is based on deviations from the periodic means, it can be called as the ‘within period estimator’. Besides the point estimator of parameters, we also want to estimate the variance. For the sake of simplicity, as well as improving efficiency, we adopt the iid assumption on error terms, and we also require unconditional homoskedasticity. Assuming that \( \{ \eta_{ikt}, i \in \mathcal{I}(t), k \in \mathcal{K}(t), t = 1 \cdots T \} \) is iid with mean zero and variance \( \sigma^2_\eta \), then \( \overline{\eta}_t \) is asymptotically normally distributed with mean zero and variance \( n_t^{-1} \sigma^2_\eta \). Denote the variance of \( \tilde{\eta}_{ikt} \) as \( \sigma^2_{\tilde{\eta}} \), then

\[ \sigma^2_\eta = \text{var}(\eta_{ikt} - \overline{\eta}_t) = \text{var}(\eta_{ikt}) + \text{var}(\overline{\eta}_t) - 2\text{cov}(\eta_{ikt}, \overline{\eta}_t) = \frac{n_t - 1}{n_t} \sigma^2_\eta. \] (9)
As \( n_t \to \infty \), \( \sigma^2_\eta \) is asymptotically equivalent to \( \sigma^2_\eta \). Note that although \( \eta_{ikt} \) is homoskedastic across \( i, k \) and \( t \) (so is \( \hat{\eta}_{ikt} \) asymptotically), \( \hat{\eta}_t \) is heteroskedastic across period because \( n_t \) is variable over time. The variance of \( \hat{\eta}_t \) can be estimated as \( \hat{\sigma}^2_\hat{\eta} = (n_t - 1)^{-1} \hat{\sigma}^2_\hat{\eta} \), where \( \hat{\sigma}^2_\hat{\eta} \) is consistently estimated by

\[
\hat{\sigma}^2_\hat{\eta} = \frac{\sum_t \sum_k \sum_i (\hat{H}_{ikt} - \hat{Z}_{ikt}'\hat{\beta})}{\sum_t n_t - T - M}.
\]

In this form, \( M \) is the dimension of \( \beta \), which equals \( 2K - 1 \). Note that \( \sum_t n_t - T - M \) is used as the denominator in equation (10), because demeaning within period loses \( T \) degrees of freedom. By means of periodic demeaning, we can now identify all parameters in equation (3), except the stock \( X_t \), the catch-stock elasticity \( \alpha_2 \) and the catchability coefficient \( q_1 \). In the next stage, we recover an estimate of the stock and catchability using structural information contained in equation (2). We then estimate the underlying logistic parameters that govern the stock dynamics.

The Second Stage

In this stage, we develop an estimation approach for the parameters in equation (2) as well as \( \alpha_2, X \) and \( q_1 \). Define \( c_t \equiv \alpha_2 \hat{X}_t + \hat{q}_1 \), which can be regarded as a stock index. This term is a linear function of the log-transformed stock. Since \( \alpha_2 \) and \( q_1 \) are constants, they scale the stock index such that changes in \( c_t \) signals changes in the underlying stock. Note that \( c_t \) is the periodic constant term in equation (6). According to the OLS first-order condition or observing equation (7), it can be consistently estimated through

\[
\hat{c}_t = \hat{H}_t - \hat{Z}_t'\hat{\beta}, \quad t = 1 \cdots T.
\]

Although we are not able to recover stock in the previous regression, an index of stock abundance could be estimated through equation (11). The consistency of the estimator \( \hat{c}_t \) relies on the large sample property, so this method is applicable to cases with a large number of vessels fishing in the same period. If \( n_t \) is small, the estimated \( \hat{c}_t \) contains significant noise. Manipulating equation (11), the estimated stock is a function of \( \hat{\alpha}_2 \) and \( \hat{q}_1 \). Let \( \hat{Y}_t \equiv \exp(\hat{H}_t - \hat{Z}_t'\hat{\beta}) \equiv \exp(\hat{c}_t) \), then the stock can now be estimated through

\[
\hat{X}_t = \left[ \frac{\exp(\hat{H}_t - \hat{Z}_t'\hat{\beta})}{\hat{q}_1} \right]^{1/\hat{\alpha}_2} = \left( \frac{\hat{Y}_t}{\hat{q}_1} \right)^{1/\hat{\alpha}_2}, \quad t = 1 \cdots T.
\]

Deriving \( \hat{X}_t \) from equation (12), it is intuitive to run a regression on the logistic growth model (2), with \( X_t \) simply replaced by \( \hat{X}_t \):

\[
\left( \frac{\hat{Y}_{t+1}}{\hat{q}_1} \right)^{1/\hat{\alpha}_2} = (1 + r) \left( \frac{\hat{Y}_t}{\hat{q}_1} \right)^{1/\hat{\alpha}_2} - \frac{r}{S} \left( \frac{\hat{Y}_t}{\hat{q}_1} \right)^{2/\hat{\alpha}_2} - C_t + \varepsilon_t, \quad t = 1 \cdots T - 1.
\]

Since \( \varepsilon_t \) is assumed to be iid normally distributed with mean zero and variance \( \sigma^2_\varepsilon \), equation (13) can be estimated through maximum likelihood estimation (MLE). A desirable property of ML
estimator is its invariance to reparameterization (Davidson and MacKinnon, 1993). Let \( \Theta^{MLE} \) be a vector of reparameterized parameters: \( \theta_1 \equiv q_1^{-1/2}, \theta_2 \equiv (1 + r)q_1^{-1/2}, \theta_3 \equiv r q_1^{-2/3}S^{-1} \), and \( \theta_4 \equiv \hat{\alpha}_2^{-1} \), then the log likelihood function is

\[
\log L = -\frac{T - 1}{2} \log (2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{T-1} \left[ \theta_1 \hat{Y}_t^{\theta_1} - \left( \theta_2 \hat{Y}_t^{\theta_2} - \theta_3 \hat{Y}_t^{2\theta_3} - C_t \right) \right]^2.
\]

(14)

The ML estimator is straightforward but may have a serious identification problem. Studying the first order condition of the term in the brackets on the RHS of equation (14), \([\hat{Y}_t^{\theta_1}, -\hat{Y}_t^{\theta_2}, \hat{Y}_t^{2\theta_3}, \theta_1 \hat{Y}_t^{\theta_1} \log(\hat{Y}_t^{\theta_1}) - \theta_2 \hat{Y}_t^{\theta_2} \log(\hat{Y}_t) + 2 \theta_3 \hat{Y}_t^{2\theta_3} \log(\hat{Y}_t)]\), the first term is very similar to the second term if the change of \( \hat{Y}_t \) is small. In addition, the last term is very close to a linear combination of the first three terms. Therefore, the first order condition matrix may suffer from singularity and causes multicollinearity problem. In our empirical study, we have limited observations (time periods) and the stock variation is not significant, which exacerbates the problem. One simplification is to restrict \( \alpha_2 \) to one. By this means, equation (13) can be linearized and estimated through least square methods. In this case, let \( s \equiv r/(\hat{q}_1 S) \), also define \( \Delta \hat{Y}_{t+1} \equiv \hat{Y}_{t+1} - \hat{Y}_t \), then

\[
\Delta \hat{Y}_{t+1} = r \hat{Y}_t - s \hat{Y}_t^2 - \hat{q}_1 C_t + \epsilon_t^* \quad t = 1 \cdots T - 1.
\]

(15)

For this model, \( \epsilon_t^* \equiv \hat{q}_1 \epsilon_t \), so heteroskedasticity should be considered. This is a linear model with respect to \( Y_t, \hat{Y}_t^2 \) and \( C_t \), where \( r, s \) and \( \hat{q}_1 \) can now be estimated through least square methods such as feasible generalized least square (FGLS). Let \( \hat{\Theta}^{LS} \) be the parameter set \( \{ \hat{r}, \hat{s}, \hat{q}_1 \} \) estimated by this method. In the empirical study, we adopt model (15) to reduce the computational burden. Once \( \alpha_2 \) and \( q_1 \) 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 or CPUE-like proxies (Wilen, 1976; Tsoa et al., 1985). The difference is that some parameters have already been estimated before regressing on the logistic model. One advantage of this estimation method is that it reduces the parameters in the nonlinear regression. Unfortunately, neither \( \hat{\Theta}^{MLE} \) nor \( \hat{\Theta}^{LS} \) is consistent in general. Let’s denote \( \hat{\Theta}_{naive} \) as the parameter estimates obtained through simply replacing \( X_t \) with \( \hat{X}_t \). Because \( \hat{X}_t \) is not the true stock information, it is estimated by another regression and contains measurement error. The error-in-variables problems could lead to inconsistent estimation (Carroll et al., 1995). Combining equation (7) and (11) shows that the measurement error originates from the stock index \( \hat{c}_t \), which contains sampling errors

\[
\hat{c}_t = c_t + \hat{Z}_t \left( \beta - \hat{\beta} \right) + \eta_t \quad t = 1 \cdots T.
\]

(16)

Let \( \omega_t \equiv \hat{Z}_t (\beta - \hat{\beta}) + \eta_t \), where \( \omega_t \) is asymptotically normally distributed with mean zero and variance \( \hat{\sigma}_c^2 \). By this notation, equation (16) is simplified to \( \hat{c}_t = c_t + \omega_t \). The mean of \( \hat{c}_t \) is estimated by equation (11). With \( \hat{\sigma}_c^2 \) and the covariance matrix estimated in the first stage, the asymptotic variance (denoted as Asy.Var) of \( \hat{c}_t \) is

\[
\text{Asy.Var}(\hat{c}_t) = \hat{Z}_t \left[ \text{Asy.Var}(\hat{\beta}) \right] \hat{Z}_t + \hat{\sigma}_c^2 \eta_t \quad t = 1 \cdots T.
\]

(17)

The true stock index \( c_t \) should be used in the second stage regression, but we are only able to observe the error component estimate \( \hat{c}_t \). Manipulating equation (16), the true stock index

6
\[ c_t = \bar{H}_t - \mathbf{Z}_t \beta - \omega_t, \text{ and denote } Y_t \equiv \exp(c_t), \text{ then the true stock is} \]

\[ X_t = \left[ \frac{\exp(\bar{H}_t - \mathbf{Z}_t \beta - \omega_t)}{q_1} \right]^{1/\alpha_2} = \left[ \frac{\hat{Y}_t}{q_1 \exp(\omega_t)} \right]^{1/\alpha_2} = \left( \frac{Y_t}{q_1} \right)^{1/\alpha_2}, \quad t = 1 \cdots T. \quad (18) \]

In this form, \( \hat{Y}_t = Y_t \exp(\omega_t) \), it is \( Y_t \) that should be used in the second stage regression, but only \( \hat{Y}_t \) is observable and is actually used. Estimators based on \( \hat{Y}_t \) and \( Y_t \) can be significantly different. This problem is common to all CPUE or CPUE-like estimators. If there are both process errors and observation errors, the parameters in the Gordon-Schaefer model cannot be consistently estimated when one type of error is ignored (De Valpine and Hastings, 2002). Even if the bias is small for the parameters, it causes serious problem when they are used to derive the steady state in the deterministic bioeconomic system (Uhler, 1980). The optimal policy in dynamic fisheries models can be a highly nonlinear function of the biological and economic parameters. As a consequence, small parameter differences are amplified in the policy rule. Even for a standard biological target, small differences in parameters can lead to large differences in the implied MSY. Unfortunately, the quantity and direction of the bias due to measurement error cannot be determined by analytical methods, thus we use a simulation method to reduce the bias.

Equation (18) shows that the inconsistency in the naive regression is caused by the multiplicative error term \( \exp(\omega_t) \). Although we do not know the realized value of \( \omega_t \) in each period, its asymptotic distribution can be derived. In this case, we can use the simulation-extrapolation method (SIMEX) to reduce bias in the naive regression. SIMEX uses simulation to generate additional data sets containing different level of measurement error, and extrapolate the bias-corrected estimate based on running naive regressions with contaminated data (Carroll et al., 1996). This method is used by Gould et al. (1999) to analyze the relationship of catch and effort, where both variables contain measurement error. Their application is different with ours since they focus on the data measurement while we focus on the measurement of the stock index. Generally, the simulation-extrapolation method is implemented in three steps.

First, generate additional data sets with increasing measurement error, that is, the variance increases according to \( (1 + \rho) \hat{\sigma}_{c_t}^2 \), where \( \hat{\sigma}_{c_t}^2 \equiv \text{Asy.Var} (\hat{c}_t) \) and \( 0 \leq \rho \leq 2 \). Since \( \hat{Y}_t = Y_t \exp(\omega_t) \), the structure of measurement error is multiplicative, which can be transformed to the additive form. Let \( \omega_{mt} \) be iid normal random variable with mean zero and variance \( \hat{\sigma}_{c_t}^2 \), then the additional data set is simulated through

\[ \hat{Y}_{mt} = \exp \left\{ \log (\hat{Y}_t) + \sqrt{\rho} \omega_{mt} \right\}, \quad t = 1 \cdots T, \quad m = 1 \cdots M. \quad (19) \]

Second, conduct regressions with simulated data on equation (14) or (15), and obtain \( M \) naive estimates \( \hat{\Theta}_m(\rho) \) for each \( \rho \). With large \( M \), the sample average of naive estimators converges to \( \Theta(\rho) \), which can be estimated through

\[ \hat{\Theta}(\rho) = M^{-1} \sum_{m=1}^{M} \hat{\Theta}_m(\rho). \quad (20) \]

Finally, fit models that regard each estimate in \( \hat{\Theta}(\rho) \) as a function of \( \rho \), and extrapolate the real estimator with \( \rho = -1 \). The variance of the estimators can be estimated by the same means. Note that \( \hat{\Theta}(0) = \hat{\Theta}_{\text{naive}}, \text{ and } \hat{\Theta}(-1) = \hat{\Theta}_{\text{SIMEX}} \). Constructed in this way, \( \hat{\Theta}_{\text{SIMEX}} \) is
an approximately consistent estimator (Cook and Stefanski, 1994). There are multiple candidate models to fit \( \Theta(\rho) \) with respect to \( \rho \). We adopt linear, quadratic, and nonlinear extrapolation methods. For the nonlinear model, \( \Theta_l(\rho) = a + b/(c + \rho) \), where \( l = 1 \cdots L \). In this form, \( L \) is the dimension of \( \Theta \), and \( a \), \( b \), and \( c \) are parameters. Different extrapolation methods may yield different results. We have no prior knowledge which performs better. Generally, linear extrapolation is the most conservative among all methods. In the empirical study, linear, quadratic and nonlinear extrapolation are all used and being compared.

AN APPLICATION: 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. The logbook data managed by the Southeast Fisheries Science Center (SEFSC) consists of three relational database. 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 gears 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. In the logbook data, there are 62 commercially harvested reef-fish fishery species, including 11 grouper species and 10 snapper species. It also contains 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 only use a subset of the data, specifically, the reef fish trips from fishing statistical zone 1 to 13.

The reef-fish fishery data is complicated because a single trip may contain multiple days, multiple areas, multiple species and multiple gears. We make the following assumptions. First, a fishing trip is defined as reef fish trip if the reef-fish catch accounts for at least 70% of the total catch. On average, target reef-fish catch accounts for 71% of total reef-fish catch and 68% of total catch. Otherwise, the reef-fish catch is regarded as bycatch. Second, the effort is defined as number of crews times trip days. Third, if a trip involves multiple gears and areas, the catch and effort is assumed to be evenly distributed among each gear and area. Fourth, zone 1 to zone 13 are aggregated to one zone, that is, using the single area assumption. Fifth, we lump all reef-fish species together. Sixth, the time step is one month, and the stock is assumed to be the same to any fishermen during this period. In addition, observations with dubious information are dropped such as crew number larger than 10 or trip days longer than 8.

The final data set contains a single area, and monthly-, individual- and gear-specific trip information. There are 142 months, 8 gear types, and 147,158 trips in total. The dominant gear type in reef fishery is handline \( (q_3) \), accounting for 70% of total gears reported. Other gear types include traps \( (q_1) \), bottom longline \( (q_2) \), electrical reel/bandit rigs \( (q_4) \), trolling \( (q_5) \), gill net \( (q_6) \), divers with spears \( (q_7) \), and divers with power heads \( (q_8) \).

The result of the first stage regression is reported in table 1. Every parameter is significant at the 95% level of confidence. Due to multicollinearity, the gear-specific catchability coefficient is estimated with traps \( (q_1) \) as a reference. The gear with the highest catchability is bottom longline \( (q_2) \), and the lowest one is divers with spears \( (q_7) \). More interestingly, we reject the null hypothesis that the Schaefer production function is true in the reef-fish fishery, that is, \( \alpha_{1k} \neq 1 \). The catch-effort elasticity is significantly less than one, which means the reef-fish fishery is subject to the law of diminishing returns. Note that the conclusion that \( \alpha_{1k} < 1 \) is true no matter whether the catch-stock elasticity is restricted to one or not, because \( \alpha_2 \) is canceled out by periodic demeaning. The first stage estimation is consistent under general conditions, so the above conclusion holds without the second stage regression, and also independent of the models of population dynamics.
Table 1: Reef-fish fishery: the first stage estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(q_2/q_1) )</td>
<td>0.95932</td>
<td>0.04878</td>
<td>19.67</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \log(q_3/q_1) )</td>
<td>-1.48403</td>
<td>0.03738</td>
<td>-39.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \log(q_4/q_1) )</td>
<td>-1.16147</td>
<td>0.04446</td>
<td>-26.12</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \log(q_5/q_1) )</td>
<td>-0.51596</td>
<td>0.04060</td>
<td>-12.71</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \log(q_6/q_1) )</td>
<td>0.93711</td>
<td>0.08648</td>
<td>10.84</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \log(q_7/q_1) )</td>
<td>-2.08837</td>
<td>0.04829</td>
<td>-43.25</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \log(q_8/q_1) )</td>
<td>-1.16342</td>
<td>0.15192</td>
<td>-7.66</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_1 )</td>
<td>0.46436</td>
<td>0.01044</td>
<td>44.5</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_2 )</td>
<td>0.31426</td>
<td>0.00867</td>
<td>36.27</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_3 )</td>
<td>0.72061</td>
<td>0.00231</td>
<td>312.51</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_4 )</td>
<td>0.67191</td>
<td>0.00734</td>
<td>91.58</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_5 )</td>
<td>0.46113</td>
<td>0.01192</td>
<td>38.70</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_6 )</td>
<td>0.79056</td>
<td>0.04992</td>
<td>15.84</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_7 )</td>
<td>0.77095</td>
<td>0.01421</td>
<td>54.27</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>( \alpha_8 )</td>
<td>0.65275</td>
<td>0.06552</td>
<td>9.96</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

\( a \) Number of observations: 147,158.
\( b \) Time periods: 142.
\( c \) Root MSE (\( \hat{\sigma}_n \)): 0.97689.

After the first stage regression, the stock index \( \alpha \) can be estimated. With the estimated stock information, we run a regression on the logistic growth model. To simply the computation, the catch-stock elasticity \( \alpha_2 \) is restricted to one. The result of the second stage estimation is reported in table 2. The naive estimator corresponds to \( \hat{\Theta}(0) \), where \( \hat{q}_1 \) is not significant and other estimates are significant at the 95% level of confidence. The SIMEX estimator corresponds to \( \hat{\Theta}(-1) \), and three extrapolation methods are compared with regard to bias correction. The linear extrapolation is the most conservative method, and \( \hat{q}_1 \) is now significant at the 90% level of confidence. The other two extrapolation methods produce estimates that are all significant at the 95% level of confidence, in which the nonlinear extrapolation is more aggressive.

Table 2: Reef-fish fishery: the second stage estimation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Naive</th>
<th>Linear</th>
<th>Quadratic</th>
<th>Nonlinear</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r )</td>
<td>0.306525260</td>
<td>0.298878525</td>
<td>0.296430415</td>
<td>0.295094111</td>
</tr>
<tr>
<td>(0.080912032)</td>
<td>(0.071603393)</td>
<td>(0.070707635)</td>
<td>(0.069550649)</td>
<td></td>
</tr>
<tr>
<td>( r q_1^{-1} S^{-1} )</td>
<td>1.056604382</td>
<td>0.882531992</td>
<td>0.854647219</td>
<td>0.836336739</td>
</tr>
<tr>
<td>(0.224374802)</td>
<td>(0.206228048)</td>
<td>(0.201579790)</td>
<td>(0.198852123)</td>
<td></td>
</tr>
<tr>
<td>( \hat{q}_1 \times 10^{-5} )</td>
<td>0.716007622</td>
<td>1.279244566</td>
<td>1.379601897</td>
<td>1.424969001</td>
</tr>
<tr>
<td>(0.812635300)</td>
<td>(0.759489289)</td>
<td>(0.718014014)</td>
<td>(0.712310911)</td>
<td></td>
</tr>
</tbody>
</table>

\( a \) Catch-stock elasticity \( \alpha_2 \) is restricted to one.
\( b \) Number of observations: 141.
\( c \) The entry in the parenthesis is the standard error.

The simulation-extrapolation process is plotted in figure 1. It shows that measurement error causes a mixed effects on the naive estimate: \( \hat{\theta}_1 = \hat{r} \) and \( \hat{\theta}_2 = r/(\hat{q}S) \) are biased up, while \( \hat{\theta}_3 = \hat{q} \) is
biased down. The bias for $\hat{\theta}_3$ is quite large, since the nonlinear SIMEX estimate doubles the naive estimate. This figure also shows that all standard errors are biased up. All parameter estimates are more significant after SIMEX correction. This is reasonable because the measurement error adds noise to the observed data and slurs the estimators.

![Figure 1: Illustration of the simulation-extrapolation](image)

Having the structural parameters identified, it provides some insights on the population dynamics. As a first step, we assume the system (1) and (2) 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, maximum sustainable yield, and maximum sustainable effort. Table 3 reports the result. It shows that the naive regression overestimates the carrying capacity, and the implied maximum sustainable yield is too high. In computing the maximum sustainable effort, we only adopt one gear type (handline). The estimated effort is also overestimated by the naive regression. Converting the MSY in table 3 to yearly data, all estimates are smaller than the collective MSY 51 million pounds set by the Reef-Fish Fishery Management Plan in 1980's (Rueter, 2006). Note that since we only have commercial fishing data, the estimated MSY is smaller than the one including recreational fishing data.

Table 3: Implied MSY in a deterministic bioeconomic system

<table>
<thead>
<tr>
<th>Variable</th>
<th>Naive</th>
<th>Linear</th>
<th>Quadratic</th>
<th>Nonlinear</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S \times 10^3$ lbs</td>
<td>40,516.90</td>
<td>26,380.46</td>
<td>25,140.97</td>
<td>24,761.33</td>
</tr>
<tr>
<td>$H^\text{MSY} \times 10^3$ lbs</td>
<td>3,104.86</td>
<td>1,971.14</td>
<td>1,863.14</td>
<td>1,826.73</td>
</tr>
<tr>
<td>$E^\text{MSY}$ (crew × day)</td>
<td>7,784,910</td>
<td>1,939,206</td>
<td>1,615,747</td>
<td>1,512,529</td>
</tr>
</tbody>
</table>

* Maximum sustainable effort is computed for handline ($q_3$) only.
CONCLUSION

This paper is an extension of the previous research on estimating the Gordon-Schaefer type model. In this paper, we generalize the production function to a less restricted Cobb-Douglas function. Although the coefficient of stock is restricted to 1, we can readily relax this assumption and adopt the maximum likelihood estimation. However, this generalization significantly increases the computational intensity. Another extension is to explicitly include the unobserved stochastic term to the model, since the assumption on the error term is critical for consistent estimation. Because the error term occurs in both the production function and the logistic growth model, the traditional estimation method is not consistent if the measurement error of the CPUE-like proxy is ignored.

A two-stage estimation method is proposed to deal with the unobserved stock information and the measurement error problem. This method is based on the heterogeneous fishing data such as the logbook data, which provides more information than the aggregated data. The ‘within period estimator’ in the first stage works very well in the Monte Carlo experiment and in the empirical study. The estimated stock index derived from the first stage regression is a good proxy to reflect the change in stock. The major concern is in the second stage because the estimated stock index is an error component and the naive estimator is biased. The simulation extrapolation method is used to reduce the bias, and the plot of $\hat{\Theta}(\rho)$ versus $\rho$ is very informative. This method performs well with small measurement error. The major advantage of the two-stage estimation method is to simplify the estimation. With the catch-stock elasticity restricted to one, we only need run two linear regressions. For a more generalized model, it reduces the parameters in the second stage, which is often nonlinear and difficult to converge.

This estimation method is tested on the simulated data. It is also applied to the reef-fish fishery in the Gulf of Mexico. The empirical study shows that the second stage estimator is biased by the measurement error of the stock index. Although the naive regression and the SIMEX estimator may only have slight differences, they imply large differences in the steady state of the dynamic system. The empirical study shows that the naive regression significantly overestimates the carrying capacity, as well as the maximum sustainable yield and effort. The fishery management based on different estimators could make big difference.

ACKNOWLEDGEMENT

This research was supported by the Saltonstall-Kennedy Program, NOAA #NAO3NMF4270086.

REFERENCES


