Environ Resource Econ (2015) 61:115–125
DOI 10.1007/s10640-013-9740-1

The Endogeneity Problem in Applied Fisheries Econometrics: A Critical Review

Daniel V. Gordon

Accepted: 29 October 2013 / Published online: 16 November 2013

Abstract The purpose of this essay is to bring attention to some serious problems that exist in econometric application of fisheries economic models. These problems in application are serious to the point of impeding our ability to do policy work. This essay will focus on two areas of econometric application; first, the violation of the fundamental exogeneity condition for applied econometrics $E(\varepsilon|X) = 0$, where ε is a stochastic error term and X is a matrix of right-hand-side explanatory variables, and second, the inappropriate use of data that is available for analysis. Both problems deal with the econometric issues of omitted and proxy variables. I will also comment on data necessary to carry out proper fisheries econometric research and policy analysis. Simulation techniques based on a known population regression equation are used to illustrate the extent of the empirical problems. Injecting real data into the simulations shows the bias that we can bring forward to the policy arena.

Keywords Fisheries economics · Omitted variables · Proxy variables

JEL Classification C1 · Q22

1 Introduction

The purpose of this essay is to bring attention to some serious problems that exist in econometric application of fisheries economic models.[1] These problems in application are serious

[1] In applied econometrics, there are many problems and complications but these are put aside in this essay to concentrate on omitted/endogenous variable problems.

This essay is based on a presentation to the Workshop In Honour of Dr. Colin Clark: Developments and Challenges in Fisheries Economics, May 15–16, 2012, Vancouver, Canada.

D. V. Gordon (✉)
Department of Economics, University of Calgary, Calgary, Canada
e-mail: dgordon@ucalgary.ca

D. V. Gordon
Department of Industrial Economics, University of Stavanger, Stavanger, Norway

to the point of impeding our ability to do policy work. I will focus on two areas of econometric application; first, the violation of the fundamental exogeneity condition for applied econometrics $E(\varepsilon|X) = 0$, where ε is a stochastic error term and X is a matrix of right-hand-side explanatory variables, and second, the inappropriate use of data that is available for analysis, which can lead also to the violation of the exogeneity condition. I want to be clear that the problems addressed in this essay are not new and well developed in the econometrics literature but all to often in application we tend to overlook the importance and consequence of endogeneity on our results. In fisheries economics, the omitted variable and measurement error problems are common causes of endogeneity issues, which can lead to inconsistent parameter estimates. I will also comment on data that should be available to carry out proper fisheries research and policy analysis. To illustrate concerns with applied fisheries econometrics, simulation techniques based on known population regression equations are used to show the extent of the empirical problem. Then, real data is used in the simulations to show the bias that we can bring forward to the policy arena.

The organization of the paper is first, in Sect. 2, to review the omitted variable problem and briefly describe three procedures for correcting the problem. In Sect. 3, I place the econometric problems within two fisheries economic examples using both a bio-economic and dual profit models. Sect. 4 simulates the bias caused by inconsistent estimates using simulated and real data. Section 5 offers final comments.

2 The Omitted Variable Problem

Lets use a simple linear catch equation to describe the statistical outcome of the omitted variable problem.[2] Assume we know the true population catch (C) regression equation[3] as a function of fishing effort (E) and skipper effect (S) specified in deviation form to remove the intercept or;

$$C_t = \beta_E E_t + \beta_S S_t + \varepsilon_t \tag{1}$$

If we are interested in measuring the true effect of a change in fishing effort on catch we would like to recover a good estimate of the population parameter β_E. By good estimate we mean that $E\left(\tilde{\beta}_E\right) = \beta_E$, which says that in a theoretical sense on average we are measuring the true value. A alternative condition for good estimates and empirically more satisfying is that in the $\lim_{t\to\infty} \tilde{\beta}_E \to \beta_E$, or empirically the more data we can bring to bare on the problem the closer we are to the true parameter. If we are using aggregate time series data we can gather information on catch levels and fishing effort but often we lack information on the skipper effect and in application we ignore this variable. In estimation Eq. (1) becomes;

$$C_t = \tilde{\beta}_E E_t + v_t, \ where \ v_t = \beta_S S_t + \varepsilon_t \tag{2}$$

Notice that ignoring the skipper effect does not mean it goes away it just shows up in the error term. With simple algebra it is easy to show that $E(\tilde{\beta}_E) = \beta_E + \beta_S \delta$, where $\delta = \frac{Cov(E_t S_t)}{Var(E_t)}$. Estimating Eq. (2) does not provide a good estimate of β_E but rather an estimate of a combined measure of fishing effort (β_E) and skipper effect (β_S). Let me push this exercise a little further. Let $\delta = 1$ and say skipper effect is as important as fishing effort in determining catch

[2] See, Wooldridge (2013) chapter 3 and Hill et al. (2011) chapter 6 for excellent discussions of the omitted variable problem.

[3] In practice we never now the true population equation and this adds to the econometric difficulties for applied researchers.

levels, then $\tilde{\beta}_E$ is twice as large as the true value β_E. Ignoring the skipper effect we take our measured results to policy makers and state that a policy of removing X-vessels from the fleet will reduce catch levels by Z%. In a narrow sense our advice will only reduce catch levels by 0.5Z% but in the bigger picture because we do not control for skipper effect Eq. (2) tells us nothing of how catch levels will change with the removal of vessels from the fleet.

One cannot overstate the seriousness of the endogeneity problem for applied researchers. As we have seen, theoretically it is easy to show the problem of omitted variables but empirically we do not know the population regression equation or the extent (δ) of the omitted variable problem. However, we do have statistical tools available to us that we hope will move us in the right direction in coming up with good estimates to take to the policy arena. The way forward depends on the empirical problem at hand, the type of endogeneity problem we think we are facing and the information and data available to bring to bear on the problem.

Continuing with the linear catch example and say, using aggregate time series data in estimation the problem is that we do not have an index of the skipper effect. One procedure to correct or at least minimize the problem is to use a proxy[4] variable in estimation in place of the skipper effect. Good proxy variables must be correlated with the omitted variable and once accounted for in the regression the true variable (skipper effect) is not correlated with other right-hand- side variables in the equation. For the case at hand, perhaps the average years of skipper experience in the fleet might be a useful proxy. Average experience may not satisfy fully the theoretical demands of a good proxy but it may be empirically useful in minimising inconsistency in the parameter of interest, in this case fishing effort.

On the other hand, if the researcher has available data on individual vessels (i) overtime (t) Eq. (1) can be rewritten as;

$$C_{it} = \beta_E E_{it} + \beta_{Si} + \varepsilon_{it} \tag{3}$$

Panel data allows us to identify the skipper effect (β_{Si}) for each vessel. In fact, if the number of vessels in the data set is small and time periods long a simple dummy variable technique will allow a consistent estimate of (β_{Si}). However, if the number of vessels is large and time periods short there are a number of estimators[5] that will allow consistent estimates of (β_E). Of course, in all cases one must be careful to correct for heteroskedasticity[6] and autocorrelation. In panel data, this is a particularly serious problem where clustered heteroskedasticity is a common problem and the unobserved fixed effect forces autocorrelation.

A common procedure in econometrics to correct the omitted variable problem is the use of instrumental variables (IV).[7] To illustrate the use of IV, consider Eq. (1) but now let catch (C) be a function of fishing effort (E) and stock abundance (S^*) or.

$$C_t = \beta_E E_t + \beta_S S_t^* + \varepsilon_t \tag{4}$$

It is likely that we do not know the true count or measure of stock abundance but rather we have an estimate[8] say

$$S_t = S_t^* + \vartheta_t \tag{5}$$

4 See, Wooldridge (2013) chapter 9 and Hill et al. (2011) chapter 10.

5 In econometrics the fixed effects estimator is commonly used but also the random effects estimator is available (Wooldridge 2002, chapter 10).

6 In all cases, heteroskedasticity must be corrected in order to generate appropriate test statistics.

7 See, Cameron and Trivedi (2010) chapter 6.

8 The International Council for the Exploration of the Sea generates stock abundance variables subject to error.

In other words, we know the value of the stock up to a random component, ϑ_t. For estimation with time series data substitute Eqs. (5) into (4) and

$$C_t = \beta_E E_t + \beta_S S_t + \omega_t, \text{ where } \omega_t = \vartheta_t + \varepsilon_t \tag{6}$$

Clearly $E(\omega_t|S_t) \neq 0$ and measurement error violates the exogeneity assumption. We can over come the problem by finding an instrument or a set of instruments, say (Z) that correlates with S_t and has the property $E(\omega_t|Z_t) = 0$. In addition and most importantly Z must not be part of the model specification in Eq. (4). In other words, Z has no direct effect on catch and only impacts catch indirectly through correlation with stock abundance. For the case at hand the instruments might be lagged catch levels or the rank order (R) of S_t.[9] Estimation is carried out in a two-step procedure; first run the regression[10]

$$S_t = \alpha_o + \alpha_1 E_t + \alpha_2 C_{t-1} + \alpha_3 R_t + \epsilon_t$$

and save the predicted values $\hat{S}_t$. In the second stage, estimate Eq. (6)[11] replacing S_t with $\hat{S}_t$. This procedure will allow consistent estimates of the parameters for both fishing effort and stock abundance (β_e *and* β_S).

3 Inconsistent Estimates

In fisheries economics, we have a vast arsenal of structural models that applied to the fishery provide the theoretical foundation from which empirical modelling and policy analysis can be carried out (Clark 1976). As well, econometric theorists have provided us with a vast arsenal of estimators and techniques for testing, validation and forecasting in generating good applied econometric models.[12] It is the responsibility of applied fisheries economists to mix these two academic fields with real data to produce solid and accountable empirical work and useful policy analysis. It is the interchange of this core of work where I see two serious problems; first, is the inappropriate application of econometric techniques and, second, the misuse of available data, notwithstanding all the constraints, to carryout responsible applied work.

Examples will help illustrate the issues. It is common in fisheries economics research to estimate bio-economic models. These models as we know, combine both economic and biological structure to model the fishery. Ignore modelling the biological side other than to say that it is usually quite complicated, including year classes, stochastic behaviour etc. Lets focus on the economic model, which is based on some production function[13] leading to a profit function. The production function of choice for bio-economic modellers,[14] the *Cobb–Douglas* functional form commonly written as:

$$H_t = a_o E_t^{\beta_E} S_t^{\beta_S} e^{\varepsilon_t} \tag{7}$$

[9] See Ekerhovd and Gordon (2013) for an example of IV used to correct measurement error.

[10] A test of the validity of the IV is the joint test $H_o: \alpha_2 = \alpha_3 = 0$. A rule of thumb is that the F-statistic should be greater than 10 (Stock and Watson 2007, chapter 12).

[11] The easiest way to perform the IV estimator correctly is to use programmed 2SLS estimator, however, if manually estimating the two equations always bootstrap the second equation to recover efficient errors.

[12] See for example Wooldridge (2002), Greene (2003), Davidson and MacKinnon (2004), and for applied econometrics with Stata software see, Cameron and Trivedi (2010).

[13] Zeller et al. (1966) describe a stochastic procedure based on expected profit maximization that allows consistent estimates of the parameters of the harvest function.

[14] See, for example Bjørndal (1987).

where H_t is harvest in period t, E_t is some index of fishing effort in period t, S_t is stock abundance in period t and ε_t is some stochastic error term that includes a true random component plus *all other* factors that impact harvest but not explicitly modelled. The *Cobb–Douglas* is chosen not because it is a good approximation to the true primal function but because it is simple, requires only basic data on the fishery and the parameters are easy to interpret. The very restrictive nature on marginal rates of substitution is ignored, as is the fact that the beauty and complexity of the economic production process with factor substitutability, economies of scale, subadditivity, joint production, technological change and more is lost in some aggregate approximation index E. What is more, as we will see next even $E$is eliminated from further analysis. Continuing on, the profit function (π_t) is simply:

$$\pi_t = P_o \cdot H_t - \mathrm{c} \cdot E_t \tag{8}$$

where P_o is output price and c is constant marginal cost. Combining Eqs. (7) and (8) and taking expectations we get

$$E(\pi_t) = P_o \cdot H_t - c\left[H_t/a_o S_t^{\beta_S}\right]^{\frac{1}{\beta^E}} \tag{9}$$

There is not much economics left in Eq. (9), fishing effort (E) has been factored out and expected profits are a function only of harvest and stock levels. In fact the only economics in the equation comes from a profit maximization assumption. Although not directly related to the topic of this essay, Eq. (9) should be questioned and evaluated as to its ability to adequately reflect and model the economics in the fishery. Notwithstanding, the parameters of Eq. (9) can be recovered from econometric estimation of the harvest function, Eq. (7). The parameters of Eq. (7) can be identified and estimated if the following condition is satisfied;

$$E\left(\varepsilon_t | E_t, S_t\right) = 0 \tag{10}$$

What this says is that there must be no contemporaneous relationship between the error term and the explanatory variables, fishing effort and stock level. Assume this condition holds for the relationship between the error term and stock level but investigate the relationship between the error term and fishing effort. Fishing effort is a function of the factor inputs used in the harvesting process (Squires 1988) and in general we can write this as;

$$E_t = f(L_t, K_t, F_t, \tau_t) \tag{11}$$

where labour (L_t), capital (K_t), fuel (F_t) and some stochastic term (τ_t) (that includes a random error term and all other factors that impact fishing effort but not modelled) combine to produce fishing effort. In practise we do not observe an index of fishing effort and in fact it is commonly proxied by some index of labour or capital (Bjørndal and Conrad 1987). Using capital as a linear proxy for fishing effort we can write;

$$E_t = \alpha_o + \alpha_K K_t + \epsilon_t$$

where ϵ_t includes a random term plus all other factor inputs that impact fishing effort. Substituting the effort equation into Eq. (7) the harvest function can be written as;

$$H_t = a_o K_t^{\gamma_E} S_t^{\beta_S} e^{v_t} \tag{12}$$

Notice that by using a proxy, Eq. (12) does not allow us to recover the impact of fishing effort (β_E) on harvest rather we recover some parameter γ_E. Moreover it is tempting to define the estimate of γ_E as the change in harvest from a one-unit change in the capital index, but this is *not correct* because the error term v_t is correlated with capital and as such γ_E is not the limiting factor of $\hat{\gamma}_E$. This is clear because υ_t contains elements of labour and other factor inputs that are correlated with capital. This is an example of a poor proxy variable used to mitigate the omitted variable problem. The conclusion is clear, estimation of Eq. (12) cannot provide an estimate of β_E or provide a consistent parameter estimate of γ_E. It may be some solace that $\hat{\beta}_s$ is a consistent estimate of the true parameter β_s if S_t is not correlated with ε_t, τ_t or υ_t.

An alternative way to model the harvest process is to model in the dual (Jensen 2002). Rather than estimate a harvest function directly we can indirectly recover the parameters using a dual profit (π_t) function. If we are operating within a competitive environment, the profit function avoids many of the econometric problems in direct estimation. The dual profit function is written as;

$$\pi_t = f(P_t^o, P_t^K, P_t^L, P_t^F, S_t, \omega_t) \tag{13}$$

where P_t^o is output price, $P_t^i, i = K, L, F$ is the factor input price and ω_t is a stochastic error term. Of course, following standard duality theory (Diewert 1974) output supply and input demands are easily defined, and estimation is straightforward in generating consistent estimates given that Eq. (13) is the true equation and that the following assumption is satisfied;

$$E(\omega_t | P_t^o, P_t^K, P_t^L, P_t^F, S_t) = 0 \tag{14}$$

In words, if prices are set in a competitive market and stock levels are not correlated with the error term estimation of Eq. (13) will generate consistent estimates of the parameters. Equation (13) is more demanding in terms of data than the simple harvest function in Eq. (7) but it is also rich in characterizing the economic structure of the harvest process.

In practise problems can arise if market prices are not available. If for instance fuel prices are not available the applied researcher can either ignore this variable and it ends up in the error term causing problems as discussed above or to proxy the variable with data available. It is here that we must be careful to avoid a serious pitfall in the use of proxy variables. In the fisheries literature it is a common problem that fuel expenditures rather than prices are available.[15] In such cases, efforts are made to derive a price proxy from expenditures say:

$$\tilde{P}_t^F = P_t^F \cdot Q_t^F / K_t \tag{15}$$

or the cost of fuel per unit of capital (Asche et al. 2009). However, plugging the fuel price proxy into Eq. (13) fails to generate consistent estimates, as quantity of fuel is a choice variable and correlated with the error term. What is most important and sometimes overlooked in practise is that if the right-hand-side variables in Eq. (13) are correlated (and they most certainly are) the poor proxy not only generates inconsistent estimates of the fuel price coefficient but of *all* parameter estimates.

Whether we estimate a production function or profit function it is necessary to carry out proper validation and testing of the estimated equation. Where the applied researcher can go wrong is to try to identify and validate off the *empirics*; apply some variant of least squares and hope the signs (or at least some of the signs) are consistent with priors and small p-values, and this alone justifies the exercise. Such strategies should be avoided and will more than likely generate inconsistent parameter estimates and inappropriate policy conclusions.

There is no magic in doing good econometrics but there is good research strategy to be followed. The fundamental step in good applied research strategy is to first, write down the population/theoretical regression equation. This equation is based on a solid understanding of the economic structure of the problem, a solid understanding of the market and practical

[15] This is a typical problem of statistical agencies collecting data for accounting and not economic purposes.

aspects of the problem and most important good common sense on what you are trying to model.[16] The purpose in writing down the population regression equation is to attempt as best as possible to identify all factors that impact the dependent variable and not just the variables for data available. Success at this stage does not eliminate econometric problems but it does allow you to identify problems. How often have we heard young academics (and perhaps not so young) comment that the regression results are good except one parameter has the wrong sign! This is a clear flag that all regression results should be considered suspicious and one should ask the question what is in the error term?

4 Simulations

Lets see why this is such a serious issue. Write the population regression as;

$$H = f(W, \varepsilon) = f(X, Z, \varepsilon) \tag{16}$$

the vector W is made up of all population variables that impact harvest. You can think of W as a combination of a statistically known vector of variables X and statistically unknown or unobserved vector of variables Z. We cannot estimate the parameters of Eq. (16) because we don't know the vector Z, so we write down a sample regression equation;

$$H_t = g(X_t, v_t) \tag{17}$$

where $v_t = k(Z_t, \epsilon_t)$, ϵ_t is the sampling error. The important point here is that the unknown and unobserved vector of variables Z_t does not vanish in estimation but is part of and impacts the sample regression equation through the error term. Serious problems arise in estimation and interpretation (inconsistent estimates) if the vectors Z_t and X_t are correlated.

An example will help clarify the issue. The variables in the population harvest Eq. (16) might include labour, capital, fuel, other factor inputs, stock abundance, environmental factors, management, technical change and unobserved factors. The sample harvest Eq. (17) might only include labour, capital, fuel and stock abundance. The whole point of writing down the population regression equation is to help the researcher define what is in the residual error term in the sample regression equation or;

$$v_t = \kappa\,(Z, \epsilon_t) \tag{18}$$

Now, it is easy to see that the residual error is correlated with the vector X (i.e., factors in X are correlated with factors in Z) and least squares applied to the sample harvest equation will produce biased and inconsistent parameter estimates. Keep in mind that the correlation problem cannot only change the magnitude of estimated coefficients, but also the standard error and possibly the sign of the coefficient.

Now we need to address two concerns; first how do we avoid this serious econometric problem and second, surely the least squares approximation of Eq. (17) is adequate to provide reasonable policy analysis? For the former this is the art of model building. What is required is to include sufficient direct and control variables[17] that allow the researcher to make a causal interpretation of the estimated parameters. What this means is that you want to specify the regression equation to such an extent that the estimated beta coefficient does in fact reflect the impact of a change in an independent variable on the dependent variable holding all

[16] In fact, the applied researcher will never know the true population regression equation.

[17] Where direct variables are the variables of direct interest to the study and control variables are to control the correlation and variance of the estimates.

other variables constant. This will include the appropriate use of proxy, instrumental and lagged variables. This of course is difficult and in many cases may not be possible. In order to proceed, it is up to the researcher to argue the possible direction of the bias in estimation. Even with bias or inconsistent results, estimation and interpretation may still be possible if the bias is in the direction of rejection of prior expectations i.e., in favour of the Null hypothesis.

For the latter, we can use simulation techniques to show how serious the bias in estimation and policy work can be with a least squares estimator. Let us first use simulation techniques to estimate the parameters of the harvest function using a least squares estimator under Gauss–Markov conditions (i.e., $E(\epsilon_t|E_t, S_t) = 0$).[18] Assume we know the harvest population regression equation[19] and it is;

$$H_t = 10 \cdot E_t^{2.0} \cdot S_t^{4.0} e^{\epsilon_t} \tag{19}$$

For our exercise we will simulate residual error ϵ_t as $\epsilon_t \sim N(0, 1)$. Effort is simulated using a Chi-square random variable with three degrees of freedom and stock is simulated as a random normal variable around a deterministic trend. Sample size is set at 10,000 observations and using the simulated values for effort, stock and residual, Eq. (19) recovers the harvest level. Next, ignore Eq. (19) and use the values for harvest, effort and stock in a least squares regression to recover the parameters of the harvest equation. We report the estimated coefficients under Gauss Markov in Table 1. What does this tell us? If we know the true population regression equation and Gauss–Markov is satisfied, then least squares estimation of the sample regression equation can recover the true parameters of the population equation. Of course, econometric theory told us this would be the outcome but it is always nice to empirically verify the truth.

Now, lets simulate some aspects of empirical reality. The applied researcher does not know the exact form of the population equation and uses say vessel size to proxy fishing effort. As we saw in Sect. 3 the proxy will be correlated with the sample error term.[20] Assume for now that stock level is not correlated with fishing effort or the sample error term. In estimation we use the harvest level and stock size simulated from the population regression equation, combine this with vessel size as the proxy for fishing effort. The least squares results are reported under row headed endogeneity of Table 1. First note, as we described above that even with endogeneity between the proxy for fishing effort and the error term, least squares provides a consistent estimate of the coefficient of stock size on harvest level. However, the correlation between vessel size and error term causes a positive bias in the estimated parameter of fishing effort on harvest level. Under these simulation conditions the bias is about 12 % of the true population parameter and, what is more, the applied researcher would take to the policy table an argument that the impact of fishing effort on harvest level is 12 % higher than the actual value. (Using the estimated results a null hypothesis that the true value is 2.0 is easily rejected.)

Now lets ask what happens if stock level is correlated with fishing effort but maintain the no correlation between stock level and the error term. We re-estimate the sample regression under the new assumptions and report the results under endogeneity and correlation in

[18] Of course, under Gauss–Markov (as theory dictates) we will generate unbiased and consistent parameter estimates.

[19] Of course, in practise we never know the true population regression and this is the core of applied econometrics.

[20] Vessel size is simulated as the effort variable correlated with the residual. Of course, the extent of the correlation between vessel size and the error term will impact the degree of parameter inconsistency. Our purpose here is not to put a value on the bias but rather to show that under very conservative correlation conditions the inconsistency is a serious policy concern.

Table 1 Simulation harvest function

	Fishing effort	Stock level
Gauss–Markov	2.000	4.000
	(0.000)[a]	(0.000)
Endogeneity	2.249	3.999
	(0.008)	(0.000)
Endogeneity and correlation	2.279	3.925
	(0.000)	(0.000)

[a] p value reported in parentheses

Table 1. Notice that the bias in the coefficient for fishing effort is maintained (slightly higher) under the new conditions but the coefficient for stock size is biased downward.[21] In words, even with stock level uncorrelated with the error term in practical application the correlation of fishing effort and error term statistically impacts the estimated parameter on stock level.

Let us take the simulation exercise one step further and use actual data for a fishery [the Scotia-Fundy ground fishery off the east coast of Canada, (Dupont and Gordon 2007)] to simulate the extent of the bias that the researcher is likely to confront in applied work. We will use a dual profit function approach to show the impact of an inappropriate proxy variable in practical use. Again, assume we know the population equation and define the profit function as;

$$\pi_t = 326.0 \cdot P_{ot}^{0.9} \cdot P_{ft}^{-0.4} \cdot L_t^{0.3} \cdot K_t^{1.2} \cdot e^{\varepsilon_t} \tag{20}$$

Labour and capital are assumed fixed in the short run that is why they appear as quantity variables and ε_t is i.i.d. and $\varepsilon_t \sim N(0, 1)$. Based on the known population profit function [(Eq. (20)] and using actual data (observations 101) for the Scotia-Fundy ground fishery we can generate profit values. We will use a Monte Carlo procedure and repeat the exercise 1,000 times. We start my reporting least squares Gauss–Markov results in column 2 of Table 2. Again our results are consistent with theory and knowing the population regression coefficients under Gauss–Markov the least squares simulation recovers the true values.

Next, assume that the price of fuel is not available to the researcher but fuel expenditure is available. Lets use Eq. (15) to build a proxy variable for price of fuel. Remember the profit variable is generated from the true population equation, the proxy defines the price of fuel and all other variables are actual data for the fishery. We report the average results for the Monte Carlo experiment in column 3 headed fuel proxy in Table 2. Notice that the coefficients for labour and capital are little changed with the use of the proxy. On the other hand, the coefficient on price of output is now recorded as statistically significant and negative in sign and although the coefficient on fuel proxy is negative it is statistically insignificant. With the benefit of knowing the population regression equation we are aware that the proxy is completely inappropriate. However, for the applied researcher who has not thought carefully about the population regression equation or the development of the proxy there is a tendency to identify off the *empirics*. Such a researcher will be driven to find other data that can be added to the sample regression to make corrections, i.e., obtain statistical results consistent with priors. For the case at hand, it turns out that the actual data is a cross-section of vessels based on 3 years of data. If we statistically forge ahead, lets introduce year dummy variables to the regression even though they are not part of or specified in the population regression

[21] In our conservative exercise, the stock parameter is biased downward by about 1 % and statically important.

Table 2 Monte Carlo profit function

	Gauss–Markov	Fuel proxy	Data mining
P_o	0.896	−5.187	0.7709
	(0.322)[a]	(0.477)	(0.454)
P_F	−0.399	−0.006	−0.002
	(0.029)	(0.005)	(0.001)
Labour	0.300	0.300	0.299
	(0.001)	(0.001)	(0.001)
Capital	1.199	1.189	1.190
	(0.001)	(0.001)	(0.053)
D90	–	–	−1.274
			(0.262)
D91	–	–	−3.795
			(0.344)

[a]Standard error reported in parentheses

equation[22] and re-estimate the results. These results are reported in column four headed data mining in Table 2. Notice the dummy variables are statistically important and the coefficient on the price of output is now positive and almost statistically important (t =1.69). (Keep in mind that the empirical dummy results are spurious and do not belong in the sample regression regression. I use this example to show how available data can drive research strategy, even though the results are spurious.) What is more, the dummies have maintained the sign and lowered the standard error on the proxy fuel coefficient to statistically important (t = 1.818). Driven by the empirics and with priors satisfied the researcher now concludes that the sample regression equation has uncovered the true relationship and moves to policy implications. Certainly from an empirical point of view the final estimated equation looks reasonably sound with only minor statistically significant problems in coefficients of price of output and the fuel proxy. But clearly the research strategy followed has lead to a spurious relationship.

What about policy analysis? The estimated results report a fuel price coefficient that is less than 1 % of its true value and this means that the researcher would argue for a more inelastic demand for fuel than what is true. Moreover, the dummy variables imply a serious decline in profit over the 3-year period, when in fact no such decline occurred. In other words using empirics to guide the research results in seriously inappropriate policy conclusions.

5 The Way Forward

One might get the impression that applied econometrics is too difficult and too uncertain to lead to useful results and policy advice but this is incorrect. The fact is not that the field of econometrics is lacking but rather the failure of our discipline to hold researchers accountable. The problems addressed in this essay are not new; the econometrics literature has addressed these issues and offered numerous techniques to identify and to recover consistent parameter estimates. What is required is for editors and referees of fisheries economic journals to

[22] For the sceptic, introducing dummy variables to the original Gauss–Markov regression generated D90 = 0.987 with standard error 1.834 and for D91 = 2.421 with standard error 5.492. Of course, as irrelevant variables in the true population regression they are statistically unimportant in the sample regression.

demand of authors a section in their papers that argues the ‘identification’ of all econometric equations estimated. (This requirement must also extend to research agencies like the FAO.) Making this mandatory will result in many papers rejected and removed from the public domain but more importantly it will allow our discipline to regain credibility within the economics profession and policy arena.

Finally, all the fascinating techniques and power of econometrics are for not if the appropriate data are not available to address the important problems and issues. Of course, collecting appropriate economics data to address current and future fisheries problems takes time but the problem can be dealt with by our international associations; IIFET and NAAFE, where committees of econometricians and fisheries economists identify data needs and pressure governments and international agencies (i.e., FAO) to collect as required. The problems addressed in this essay are important and can be corrected if the discipline is willing.

Acknowledgments I would like to thank Yutaro Sakai and two anonymous referees for comments and suggestions that improved the style and content of the paper.

References

Asche F, Bjørndal T, Gordon DV (2009) Rents in an individual quota fishery. Land Econ 85(2):280–292
Bjørndal T (1987) Production economics and optimal stock size in a North Atlantic Fishery. Scand J Econ 89:145–164
Bjørndal T, Conrad JM (1987) The dynamics of an open access fishery. Can J Econ 20:74–85
Cameron AC, Trivedi PK (2010) Microeconometrics using stata. Stata Press, College Station, TX
Clark CW (1976) Mathematical bioeconomics. Wiley, New York
Davidson R, MacKinnon JG (2004) Econometric theory and methods. Oxford University Press, Oxford
Diewert WE (1974) Applications of duality theory. In: Intriligator MD, Kendrick DA (eds) Frontiers of quantitative economics, vol 2. North-Holland, Amsterdam
Dupont D, Gordon DV (2007) Shadow prices for fishing quota: fishing with econometrics. In: Bjørndal T, Gordon DV, Arnason R, Sumaila UR (eds) Advances in the economics of the fishery. Blackwell, Oxford, pp 87–106
Ekerhovd NA, Gordon DV (2013) Catch, stock elasticity and an implicit index of fishing effort. Mar Resour Econ (forthcoming)
Greene WH (2003) Econometric analysis. Prentice Hall, Upper Saddle River, NJ
Hill RC, Griffiths WE, Lim GC (2011) Principles of econometrics, 4th edn. Wiley, New York
ICES (2012) Report of the Arctic fisheries working group. ICES CM 2012/05:ACOM
Jensen CL (2002) Application of dual theory in fisheries: a survey. Mar Resour Econ 17:309–334
Manski CF (2007) Identification for prediction and decision. Harvard University Press, Cambridge
Squires D (1988) Production technology, costs, and multiproduct industry structure: an application of the long-run profit function to New England fisheries. Can J Econ 21(2):359–378
Stock JH, Watson MW (2007) Introduction to econometrics. PearsonAddison Wesley, Boston
Wooldridge JM (2002) Econometric analysis of cross section and panel data. MIT Press, Cambridge
Wooldridge JM (2013) Introductory econometrics, 5th edn. South-Western Cengage Learning, Mason
Zeller AJ, Kmenta J, Dreze J (1966) Specification and estimation of Cobb–Douglas production function models. Econometrica 4(34):784–795