Marine Resource Economics, Volume 14, pp. 215–223
Printed in the U.S.A. All rights reserved

0738-1360/99 $3.00 + .00

A Technical Note on Input Price Proxies Used in Salmon Farming Industry Studies

STEIN ØSTBYE
University of Tromsø

Abstract *A frequently used data set on Norwegian fish farms has until recently not contained any data on feed price. Previous work has been based on a proxy variable approach that is not efficient. In this note, a more efficient approach is suggested. An ex post analysis, using data on feed price which is now available, is carried out to discriminate between alternative approaches.*

Key words *Proxy variable methods, replication, restricted cost function, salmon farming.*

Introduction

In the last decade, there have been many studies based on a renowned data set on Norwegian fish farms, annually updated by the Norwegian Directorate of Fisheries. Some of the studies have been published internationally, and two have appeared in this journal (Salvanes 1989; 1993).[1] A problem encountered in all studies using data prior to 1994 is missing feed price data. Since feed is by far the dominant variable input in farming, this is a major concern. The solution has been to use feed expenses divided by output as a proxy variable. On the assumption of fixed proportions between output and feed input, the proxy takes the same value as the feed price up to a (multiplicative) constant. This kind of approach might have been necessary, but the question is to what extent the results are sensitive to this approximation. After all, if the proxy takes on a high value for a particular farm, the reason might be that the input-output ratio is large because the farm is badly managed rather than that the input price is high.

If the approach or conclusions based on the results pretend to have any generality beyond the sample, it is useful to replicate the studies when better information becomes available.[2] In the present paper, data for 1994 and 1995, comprised of information on feed price, are used to investigate this question. It will be shown that the basis of the proxy that is currently used is inconsistent. A new, more efficient proxy variable is suggested. The performance of the old and the new proxy is compared, using the data set for 1994 and 1995.

The paper is organized in four sections, following this introduction. They contain the conceptual framework, data description, estimation issues and results, and finally, conclusions.

Stein Østbye is an associate professor at the Norwegian College of Fishery Science at University of Tromsø, N-9037, Tromsø, Norway, e-mail: steino@nfh.uit.no.

I want to thank colleagues at the Norwegian College of Fishery Science in Tromsø, notably Terje Vassdal and Bergur Agustsson, for introducing me to the field and giving me access to the data used in this paper. I am also grateful for comments provided by two anonymous referees and the editor of this journal.

[1] Other studies published in international refereed journals include Bjørndal and Salvanes (1995).
[2] Replication and re-analysis is discussed in Arulampalam *et al.* (1997).

Framework

If the proxy variable approach works well, it should do so irrespective of the sample period. Hence, it should not matter whether we consider the sample from 1982 and 1983 used in Salvanes (1989) and (1993), or the sample from 1994 and 1995. Since the latter sample is comprised of information on feed price, we can empirically test how well the proxy variable approach works. If the performance is poor for the 94–95 sample, there is no reason why it should perform better for any other sample, including 82–83.

The simple conceptual framework presented in Salvanes (1993) is used as a starting point. I would like to emphasize that this work is confined to the proxy variable issue and that the problem of optimal model selection, in general, is not considered. In this sense, the investigation is partial, and the results are conditional on the validity of other parts of the framework that are not investigated. I assume readers are familiar with duality theory at the textbook level. If not, textbooks, such as those by Varian (1992), Chambers (1988), and Berndt (1991), might be useful.

The farm is assumed to be a price taker in the markets for feed and labor, minimizing costs conditional on licensed capacity. Smolts are assumed to be weakly separable from the inputs mentioned. Imposing symmetry, as implied by Young's theorem, and using the popular translog specification, we can write the restricted variable cost function for the farm, summarizing economic behavior and technology, as

$$v = \beta_0 + \mathbf{x}\beta + \mathbf{x}\mathrm{B}\mathbf{x}'. \tag{1}$$

The matrix form is convenient in order to economize on notation. The v stands for the log of variable costs, and β_0 is a constant term. The logs of the wage rate (w), feed price (f), output (y), and capacity (k) are elements in the vector $\mathbf{x}$, and β is a column vector of corresponding parameters $[\beta_w \beta_f \beta_y \beta_k]'$. The third term in equation (1) represents second order and interaction effects, with B as an upper triangular parameter matrix,

$$\begin{bmatrix} \frac{1}{2}\beta_{ww} & \beta_{wf} & \beta_{wy} & \beta_{wk} \\ 0 & \frac{1}{2}\beta_{ff} & \beta_{fy} & \beta_{fk} \\ 0 & 0 & \frac{1}{2}\beta_{yy} & \beta_{yk} \\ 0 & 0 & 0 & \frac{1}{2}\beta_{kk} \end{bmatrix}.$$

The double subscript refers to what second order or interaction term the parameter is assigned to; *e.g.*, β_{ff} is assigned to the second order feed price term and β_{fy} to the interaction term between feed price and output. Invoking Shephard's lemma, we obtain the cost share for feed (S_f) from equation (1):

$$S_f \equiv \frac{\partial}{\partial f}(\mathbf{x}\beta + \mathbf{x}\mathrm{B}\mathbf{x}') = \beta_f + \beta_{ff} f + \beta_{wf} w + \beta_{fy} y + \beta_{fk} k. \tag{2}$$

Since $S_f + S_w \equiv 1$ we can retain the share equation for labor if we know β_f, β_{ff}, β_{wf}, β_{fy}, and β_{fk}. We also use the cost flexibility, S_y.

$$S_y \equiv \frac{\partial}{\partial y}(\mathbf{x}\beta + \mathbf{xBx'}) = \beta_y + \beta_{fy} f + \beta_{wy} w + \beta_{yy} y + \beta_{ky} k. \tag{3}$$

If we had assumed profit maximizing, S_y would be equal to revenues over costs. Now, we can generally decompose any cost share into input quantity, input price, and variable costs and write specifically the feed cost share as $\log(S_f) = m + f - v$, where m is (the log of) feed quantity. Differentiating logarithmically with respect to y, we get

$$\frac{1}{S_f}\frac{\partial S_f}{\partial y} = \frac{\partial m}{\partial y} + \frac{\partial f}{\partial y} - \frac{\partial v}{\partial y}. \tag{4}$$

Price taking in the feed market implies that the second term to the right is 0, $\partial v/\partial y$ is the cost flexibility, and $\partial S_f/\partial y$ is obviously equal to β_{fy} [using equation (2)]. Denoting $\partial m/\partial y$ by α, we may summarize the preceding discussion, replacing equation (4) by

$$\alpha = S_y - \frac{\beta_{fy}}{S_f}. \tag{5}$$

The assumption of fixed proportions between feed input and output underlying the proxy variable approach implies that $\alpha = 1$.

Let us write the feed price proxy as $\hat{f} = m + f - \alpha y$. This is slightly more general than the usual formulation, $\hat{f} = m + f - y$. However, when $\alpha = 1$ the two are identical. We see that $\hat{f} = f$, except for a constant under fixed proportions, since $\alpha = 1$ and $m - y$ is constant. In other words, the feed price proxy is a perfect proxy close to a constant.

If the assumption of fixed proportions between feed input and output were believed to be true, the restriction implied by equation (5) with $\alpha = 1$ should be imposed on the system (1) and (2) in order to take advantage of all information available. Adding stochastic disturbance terms to equations (1) and (2), the system might be efficiently estimated by Zellner's iterative seemingly unrelated regression method (SUR), or asymptotically equivalent by three-stage least squares (3SLS), or full information maximum likelihood (FIML), depending upon what assumptions are made about the stochastic properties. If we did not take the unity restriction on equation (5) on faith, we could estimate the unrestricted model and test the restriction empirically, substituting fitted values, $\hat{S}_y$ and $\hat{S}_f$, for S_y and S_f, and the parameter estimate $\hat{\beta}_{fy}$ for β_{fy}, to obtain $\hat{\alpha}$. Now, if $\hat{\alpha}$ turns out to be significantly different from unity, the proxy could still follow the feed price close to a constant if $m - \alpha y$ is constant instead of $m - y$. If we use the familiar modelling trick of dividing all exogenous variables by their sample mean prior to taking logs and estimating the model, and evaluate α at the mean, S_y and S_f are simply equal to β_y and β_f. We can, therefore, construct a more efficient proxy variable by substituting $m + f - \alpha y$ for the feed price where $\alpha = \beta_y - (\beta_{fy}/\beta_f)$. Replacing $\alpha = 1$ by $\alpha = \hat{\alpha}$, we may revise the proxy, reestimate the unrestricted model, obtain a new estimate of α, and test if it is significantly different from the previous estimate. This iterative procedure comes to an end when convergence is obtained. We now have two alternative proxy variable candidates that can be compared to the true variable in an *ex post* analysis.

Data

The sample consists of data on 249 farms, representing about two-fifths of all 1994 licenses and 330 farms representing about one-half of all 1995 licenses. Feed price is measured as annual feed cost divided by quantity. Output is defined as an index of fish production in tons, consisting of tons harvested and the change in the stock of living fish. The wage rate is measured as the wage bill divided by hours of work. The previously used proxy variable for feed price is annual feed cost divided by output. The capacity is licensed capacity in cubic meters of water volume. In order to allow a direct comparison with previous work, all variables not quoted in physical units have been deflated to 1983 values using the consumer price index from Statistics Norway (1999). To facilitate comparison, summary statistics from 1982 and 1983 are included in table 1 (Salvanes 1993). For 1994 and 1995, I have included the input-output ratio or conversion ratio; *i.e.*, feed input per unit output, since the accuracy of the proxy hinges on the assumption that this ratio is constant across farms in the sample period.

The most notable change from the early-1980s to the mid-1990s is perhaps the large increase in average production. However, from my point of view, the most interesting change is the 50% decrease in feed price as measured by the mean of the proxy and the large decrease in its standard deviation. When comparing the proxy and the true price, it is observed that the dispersion, as measured by the standard deviation, is misleadingly large when relying on the proxy even though the distribution is much more concentrated than in the first sample period. Hence, estimates will appear to be more significant in the statistical sense when using the proxy instead of the true price. This will probably be much more pronounced in the first sample period, since it is not likely that the large difference in sample variation for the proxy will be mirrored by an equally large difference in sample variation for the true variable. As far as fixed proportions between feed and output is concerned, it would be more reassuring if standard deviations of the conversion ratio had been even smaller. If the crucial assumption behind the proxy variable is not supported by the data, the information employed in order to increase efficiency in estimation is false,

Table 1
Descriptive Statistics: Sample Means and Standard Deviations

Year (sample size)	1982 (91)		1983 (110)		1994 (246)		1995 (309)	
Statistic	Mean	St. D.	Mean	St. D.	Mean	St. D.	Mean	St.D.
Output	84	81	94	108	447	262	603	338
Feed Proxy	11.50	11.10	11.60	7.40	6.34	1.33	5.43	0.99
Conversion Ratio	n.a.	n.a.	n.a.	n.a.	1.31	0.25	1.24	0.21
Feed Price	n.a.	n.a.	n.a.	n.a.	4.86	0.71	4.41	0.60
Wage Rate	53.40	14.89	61.1	27.15	86.80	32.04	88.82	34.54
Variable Costs (1,000 Nok)	1,362	1,599	1,632	1467	3,435	1,982	3,920	2,272

Note: Variables are inflated or deflated to 1983 values by the consumer price index. n.a.= not available.

and more efficient estimates *ex ante* turn out as less efficient estimates *ex post.* Note that the mean conversion ratios are larger than reported by the Directorate of Fisheries. This is because obvious outliers have been removed from the sample.

Estimation

In order to remain closer to previous work, I have employed the Seemingly Unrelated Regression (SUR) estimation procedure. This procedure is only asymptotically equivalent to Full Information Maximum Likelihood (FIML), and parameter estimates will be somewhat different in smaller samples. In the present case, differences are small. The cost function and the feed share equation have been estimated simultaneously, assuming that error terms are bivariate normally distributed with zero mean and constant covariance matrix. Symmetry and linear homogeneity in input prices have been imposed, and four different specifications have been estimated. In Model A, the traditional proxy variable has been used. Model B is Model A with the unity restriction on equation (3) imposed. In model C, the alternative proxy is used with the consistent restriction on equation (3) imposed. Model D is the benchmark model, where the «true» feed price is used.

The parameter estimates are relegated to the appendix. The more intuitive own-price elasticities and the Allen-Uzawa elasticity of substitution based on the parameter estimates are presented in table 2. Since it may be of interest to compare results with previous work, the reported elasticities from Salvanes (1993) are included in the last column.

Standard errors are computed using the Delta-method based on the Slutsky theorem (see Greene 1997, p. 278). As expected, the estimated standard errors are smaller when using a proxy variable than in the benchmark case.[3] Comparing the two sample periods, demand for feed appears to have become more inelastic, while demand for labor has become much more sensitive to relative changes in the wage rate. Although labor and feed still are substitutes, as required by the theory when only two inputs are considered, the possibility for substitution seems to have been very much reduced. Now, concentrating on the 1994–95 sample, we may compare Model A-Model C to the benchmark model. The overall picture is that the estimates

Table 2

Elasticities Evaluated at the Sample Mean: Point Estimates and (Standard Errors)

	Model A	Model B	Model C	Model D	1982–83
Own price feed	–0.031	–0.030	–0.031	–0.080	–0.48
	(0.003)	(0.003)	(0.003)	(0.010)	(0.17)
Own price labor	–8.789	–8.814	–8.648	–7.432	–1.14
	(0.142)	(0.143)	(0.138)	(0.230)	(0.44)
Substitution	0.148	0.140	0.146	0.361	1.57
	(0.016)	(0.016)	(0.016)	(0.043)	(0.56)

Note: Estimates for Models A-D are statistically significant at the 1% level.

---

[3] We would expect the standard errors for 1982–83 to be smaller than reported. This suggests that another procedure has been used and that the figures are not comparable as far as the dispersion is concerned.

are quite similar for all the models where a proxy variable is used. Starting with Model A, where the traditional proxy is used, the elasticity of feed with respect to output at the sample mean was estimated to 0.990. Although significantly different from unity (*t*-value –3.41), it is not a bad approximation to the unity elasticity consistent with the assumption behind the constructed proxy variable. Compared with the benchmark estimate of 0.935, we understand why Model A performs slightly better than the *ex ante* more efficient Model B, since the estimates are constrained in the wrong direction. Had the benchmark estimate exceeded unity, there would have been a slight improvement. Turning to Model C, the feed-output elasticity is not statistically significant from 0.87 at the 5% level, which is used for α in the construction of the alternative proxy.[4] Constraining the feed-output elasticity to equal 0.87 at the sample mean using the proxy, gives a small improvement in the performance as measured by the average error from the benchmark estimates. The improvement is due to a considerable reduction in the gap between the proxy variable estimate and the benchmark estimate of the own price elasticity for labor. Overall, the use of any of the alternative proxies exaggerates the labor demand elasticity and the lack of substitution between variable inputs.

Concluding Discussion

What can we conclude from this analysis? The first point is that efficiency gains *ex ante* can be achieved by securing internal consistency in the model. This is not done in previous work since (*i*) the fixed input-output assumption behind the proxy variable is taken on faith and not tested for and (*ii*) the assumption is not imposed on the technology. If the assumption is rejected by the data, there is inconsistency between the sample information and the model. If the assumption is not imposed, there is, in general, inconsistency between the estimated technology and the construction of the feed price proxy. The approach suggested in this paper offers a remedy to these shortcomings. Now, another question is whether the suggestion improves the performance in the light of hindsight or *ex post*. This is an open question and the answer depends on how close the estimated elasticity of feed with respect to output is when using the new proxy compared to the use of the feed price. With reference to the available information for 1994 and 1995, there is some evidence that the suggested proxy represents an improvement in terms of performance.

There is another more conventional approach to the issue we have been discussing that could have been explored. Information from 1994 and 1995 could be used to construct a third proxy variable that closely follows the true variable in the sample period. We could have made a backcast for the 1982–83 sample on the basis of the parameter estimates and information available at that time. Replicating previous studies, results based on the alternative proxy variables could then be compared to see if the results were sensitive to the respecification. If they were, the general conclusion must be that one should be cautious since the results are not robust. However, unlike the approach we have chosen, this procedure is obviously sensitive to structural change between the two sample periods, and the possibilities for testing for long-term stability are very limited.

---

[4] I have done a grid search. For trial values equal to 0.87 and lower, the difference between the value and the estimated elasticity is not statistically different. Going up to 0.88, the estimated elasticity is statistically different from 0.88 at the 5% level. We could, of course, have chosen some other level of significance as demarcation criterion if we had preferences for a specific tradeoff between type I and type II errors.

References

Arulampalam, W., J. Hartog, T. MaCurdy, and J. Theeuwes 1997. Replication and Re-analysis. *Labour Economics* 4:99–105.
Berndt, E.R. 1991. *The Practice of Econometrics: Classic and Contemporary*. Reading, MA: Addisson-Wesley.
Bjørndal, T., and K.G. Salvanes. 1995. Gains from Deregulation? An Empirical Test for Efficiency Gains in the Norwegian Aquaculture Industry. *Journal of Agricultural Economics* 46:113–26.
Chambers, R.G. 1988. *Applied Production Analysis: A Dual Approach*. New York: Cambridge University Press.
Greene, W.H. 1997. *Econometric Analysis*, 3rd ed. Upper Saddle River, NJ: Prentice-Hall.
Salvanes, K.G. 1989. The Structure of the Norwegian Fish Farming Industry: An Empirical Analysis of Economies of Scale and Substitution Possibilities. *Marine Resource Economics* 6:349–73.
___. 1993. Public Regulation and Production Factor Misallocation. A Restricted Cost Function for the Norwegian Aquaculture Industry. *Marine Resource Economics* 8:50–64.
Statistics Norway 1999. http://www.ssb.no/emner/08/02/10/kpi/1-7.txt
Varian, H.R. 1992. *Microeconomic Analysis*, 3rd ed. New York: Norton.

Appendix

Cost Function Parameter Estimates

N = 554. Variance of dependent variables: share equation 0.36161E-02, cost equation 0.28328

Model A

Traditional Proxy

Parameter	Estimate	Standard Error	t-statistic
β_f	0.825950	0.205733E-02	401.468
β_{ff}	0.122483	0.266992E-02	45.8752
β_{fy}	0.036347	0.378535E-02	9.60209
β_{fk}	–0.025192	0.586810E-02	–4.29308
β_0	15.6840	0.301686E-02	5,198.78
β_y	0.946442	0.544599E-02	173.787
β_k	0.037410	0.866479E-02	4.31748
β_{yy}	0.020210	0.696403E-02	2.90209
β_{kk}	–0.975797E-02	0.017649	–0.552882
β_{yk}	–0.399216E-02	0.887224E-02	–0.449961

Share Equation
Sum of squared residuals = 1.42100
Std. error of regression = 0.050646
Variance of residuals = 0.256499E-02
Durbin-Watson statistic = 1.47987
$\bar{R}^2$ = 0.29

Cost Equation
Sum of squared residuals = 2.37881
Std. error of regression = 0.065528
Variance of residuals = 0.429389E-02
Durbin-Watson statistic = 1.48453
$\bar{R}^2$ = 0.98

Model B

Traditional Proxy, Restricted Model

Parameter	Estimate	Standard Error	t-statistic
β_f	0.825727	0.205987E-02	400.863
β_{ff}	0.123775	0.265442E-02	46.6297
β_{fy}	0.036341	0.380766E-02	9.54425
β_{fk}	–0.025068	0.590515E-02	–4.24506
β_0	15.6839	0.305045E-02	5,141.50
β_k	0.029315	0.840017E-02	3.48977
β_{yy}	0.028948	0.663832E-02	4.36081
β_{kk}	–0.977543E-02	0.018022	–0.542407
β_{yk}	–0.595337E-02	0.904031E-02	–0.658535

Share Equation
Sum of squared residuals = 1.42634
Std. error of regression = 0.050741
Variance of residuals = 0.257462E-02
Durbin-Watson statistic = 1.48254
$\bar{R}^2$ = 0.29

Cost Equation
Sum of squared residuals = 2.43772
Std. error of regression = 0.066334
Variance of residuals = 0.440021E-02
Durbin-Watson statistic = 1.49270
$\bar{R}^2$ = 0.98

Model C

Alternative Proxy

Parameter	Estimate	Standard Error	t-statistic
β_f	0.823483	0.207818E-02	396.253
β_{ff}	0.124110	0.266637E-02	46.5463
β_{fy}	0.022926	0.375393E-02	6.10712
β_{fk}	–0.024594	0.585857E-02	–4.19804
β_0	15.6680	0.306497E-02	5,111.95
β_k	0.032413	0.838676E-02	3.86476
β_{yy}	0.013151	0.640746E-02	2.05239
β_{kk}	–0.956009E-02	0.017656	–0.541450
β_{yk}	–0.211416E-02	0.882410E-02	–0.239589

Share Equation
Sum of squared residuals = 1.41492
Std. error of regression = 0.050537
Variance of residuals = 0.255401E-02
Durbin-Watson statistic = 1.48190
$\bar{R}^2$ = 0.29

Cost Equation
Sum of squared residuals = 2.40459
Std. error of regression = 0.065882
Variance of residuals = 0.434042E-02
Durbin-Watson statistic = 1.49099
$\bar{R}^2$ = 0.99

Model D

Benchmark

Parameter	Estimate	Standard Error	t-statistic
β_f	0.819319	0.220264E-02	371.972
β_{ff}	0.094588	0.641301E-02	14.7493
β_{fy}	0.045950	0.480810E-02	9.55685
β_{fk}	–0.036204	0.728393E-02	–4.97041
β_0	15.6809	0.010620	1,476.58
β_y	0.879211	0.017099	51.4195
β_k	0.095959	0.030639	3.13187
β_{yy}	0.014600	0.042807	0.341065
β_{kk}	0.052328	0.109380	0.478406
β_{yk}	–0.049209	0.054777	–0.898341

Share Equation
Sum of squared residuals = 1.38185
Std. error of regression = 0.049943
Variance of residuals = 0.249432E-02
Durbin-Watson statistic = 1.53720
$\bar{R}^2$ = 0.31

Cost Equation
Sum of squared residuals = 15.0836
Std. error of regression = 0.165005
Variance of residuals = 0.027227
Durbin-Watson statistic = 1.92231
$\bar{R}^2$ = 0.90