Measurement and determinants of efficiency in crop production in Nepal

Satis Devkota[28]
Wayne State University

Mukti Upadhyay[29]
Eastern Illinois University

Introduction

The share of agriculture in Nepal's GDP has been falling over time. Yet, this sector still accounts for over a third of GDP and about two-thirds of total employment in the country. Unfortunately, the decline in agriculture has resulted from stagnant or declining productivity in agriculture itself, and not because manufacturing or industry has rapidly overtaken agriculture in productivity changes. How to attain a continued rise in agricultural productivity remains a concern of policy.

Most farmers in Nepal have not attained a reasonable level of technical efficiency that farmers in its neighboring countries have achieved. Farms in Nepal are typically very small which limits the use of modern farming practices and seems to perpetuate low productivity. Our study is an attempt to quantify efficiency of farmers and estimate the gap from its potential given the technology currently prevailing in Nepal.

To estimate inefficiency in agriculture, we use a stochastic frontier production function. We also use OLS to compare the results for the Cobb-Douglas and translogarithmic functions to determine which of the two provides a better representation of the data. It is also interesting to examine what these functions yield for the levels of technical inefficiencies, returns to scale, and the elasticities of output with respect to different inputs.

Most studies of agricultural productivity in Nepal rely on small samples drawn from one specific region or another within the country. In this paper, we make use of data from the Nepal Living Standard Survey (NLSSII) collected during 2003/04 by Central Bureau of Statistics (CBS) Nepal. This is a truly representative national survey in that the samples

[28] Ph.D. candidate, Wayne State University. satisdevkota2001@yahoo.com.
[29] Corresponding author; Professor, Eastern Illinois University, mpupadhyay@eiu.edu.

were drawn from all three topographical regions and all five development zones. Our dataset comprises all households which had positive numbers for crop production and crop area. The dataset contains a total of 2535 households that meet these criteria and has some details on inputs and outputs related to agricultural production. The survey also provides a range of socio-economic characteristics at the household level. What sets of characteristics are associated with greater efficiency is also of strong interest to us since this information is likely to provide clear implications for policy.

The OLS and Stochastic Frontier Models

Our basic frontier production function can be written as follows:

$$Y_i = f(X_i \cdot \beta) \cdot TE_i \quad (1)$$

where Y_i is the actual output of farmer i, $f(X_i \cdot \beta)$ is the production function where X_i is a vector of inputs used by the farmer i and β is a set of parameters to be estimated, and TE_i is the technical efficiency achieved by the farmer i and is defined as the ratio of observed output to the maximum feasible output. $TE_i = 1$ implies the i^{th} farmer lies on the frontier having achieved the maximum feasible output while $TE_i < 1$ indicates how far below the frontier the farmer is actually producing. Thus, the technical efficiency of a farmer is the ratio of observed output to the output of the most efficient farmer and lies between 0 and 1 (Coelli and Battese, 1996). Since the function is stochastic, random shocks such as a drought or flood can affect the production process. This modifies equation (1) as follows:

$$Y_i = f(X_i \cdot \beta) \cdot \exp(-u_i) \cdot \exp(v_i) \quad (2)$$

or, in logarithms,

$$\log Y_i = X_{it}\beta - u_i + v_i \quad (3)$$

where $f(X_i \cdot \beta)$ is assumed to equal $\exp(X_i \cdot \beta)$, v_i is the stochastic noise with distribution $N \sim (0, \sigma^2_v)$. The technical efficiency term TE_i has been re-expressed as e^{-u_i}, or simply $-u_i$ in logarithms. This indicates that u_i is a non-negative random variable that still reflects the inefficiency of the farmer and is assumed to be *i.i.d*: $u_i \sim |N(0, \sigma^2_u|$ truncated on the left, whereas v_i is the random influence on production. The expected value of the farmer-specific inefficiency term u_i is defined as the conditional mean of u_i given the difference between symmetric and non-symmetric terms: $\varepsilon_i = v_i - u_i$ (Jondrow et.al., 1982).

We can also calculate the relative dominance of u_i and v_i as follows:

$$\pi = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2} \tag{4}$$

Equation (4) shows that as π approaches zero, either σ_u approaches zero or σ_v approaches infinity or both, which implies that the random error v_i is the primary determinant of the composite error ε_i. Thus the difference between the observed and frontier outputs is mainly due to random factors that are beyond the control of the farmer. In this case we cannot claim that the farmer is inefficient. On the other hand, a high value of π attributes a greater role to factors that are more in the farmer's hands.

The estimation of efficiency follows a two-step process. The first step is to estimate the frontier which leads to an estimate of the technical efficiency for each household. The second step then regresses the predicted inefficiencies against a set of variables (Z_i), particularly the household characteristics that are expected to influence inefficiency. The first step uses maximum likelihood estimation while the second uses the ordinary least squares regression (Coelli and Battese 1996). The second stage regression is given by the equation below:

$$u_i = Z_i \gamma \tag{5}$$

where, as noted above, Z_i is a set of farm-specific variables that are related to technical efficiency, and γs are respective parameters to be estimated.

We start with the Cobb-Douglas function estimated with OLS and stochastic frontier methods. The stochastic frontier is obtained by setting up a log-likelihood function where the estimation procedure chooses parameters in a way that maximizes the probability that the outputs converge to those actually observed. The OLS is a simple regression of outputs on a set of inputs where we estimate the composite errors whose variance cannot be divided into the variances of u and v. The stochastic frontier analysis, however, allows one to observe the size of farmer-specific inefficiencies separately from the random shocks.

Our variables in the production functions are measured in quantities per unit of labor used, where labor equals the sum of family, hired and exchange labor in man-days, and family labor adjusts child labor for adult equivalence.

Croplbr is the value of crops produced per unit of labor used,

Arealbr is the amount of cultivated land for crop production per unit of labor,
Fertlbr is the amount of fertilizers used per unit of labor, and
Pestlbr is the amount of pesticides used per unit of labor,

The survey data do not give a direct measure of capital input used. Most farms in Nepal do not use modern machinery such as tractors, nevertheless an omission of this input is a limitation of the present study. Second, about 66 percent of farmers in the sample use fertilizers but only about 16 percent use pesticides. Note that almost all the pesticide users (98 percent) use chemical fertilizers as well but only about 24 percent of fertilizer users also use pesticides. Further, pesticides used equal only Rs.453 among the users which amounts to barely Rs.74 for all farmers in the sample. Thus, while we note the results for pesticides, we focus more on the results for the cultivated land area and the use of fertilizers below.

Results

The OLS estimates of Cobb-Douglas function for crop value per unit of labor are as follows:

$$\begin{aligned} \ln crpval_i &= \underset{(0.081)}{7.36^{***}} + \underset{(0.018)}{0.659}\ln area_i^{***} + \underset{(0.008)}{0.147} fert_i^{***} + \varepsilon_i \\ \bar{R}^2 &= 0.462, \quad F_{2,2532} = 1090.5, \quad N = 2535 \end{aligned} \tag{6}$$

where the numbers in parentheses indicate the standard errors. All the coefficients are highly significant at one percent level. The elasticity of output per worker with respect to area under land is 0.66 and the elasticity with respect to fertilizers is 0.15. When *fert* in equation (6) is replaced with the sum of fertilizers and pesticides, the R^2 falls somewhat and causes a marginal reduction in the sum of the two elasticities (from 0.81 to 0.79). The reason is an increase in the coefficient of land-labor ratio (to 0.70) which is overcompensated by a reduction in the coefficient of other inputs (to 0.09). Thus, the inclusion of pesticides in the regression brings no gain in the efficiency of estimates.

Our translog production function has the following results:

$$
\begin{aligned}
\widehat{\ln crpval_i} = {} & \underset{(0.182)}{6.96^{***}} + \underset{(0.088)}{0.520}\ln area_i^{***} + \underset{(0.037)}{0.204}\ln fert_i^{***} + \underset{(0.011)}{.0008}(\ln area)^2 \\
& + \underset{(0.003)}{0.0411}(\ln fert)^{2***} - \underset{(0.009)}{0.0458}(\ln area * \ln fert)^{***}
\end{aligned} \tag{7}
$$

$$\bar{R}^2 = 0.491, \quad F = 489.0, \quad N = 2535$$

where the numbers in parentheses indicate the standard errors. The output elasticity values are now a function of the inputs. However, at the mean values of the inputs, the elasticities turn out almost identical: 0.65 with respect to land and 0.16 with respect to fertilizers, as compared to 0.66 and 0.15 respectively under Cobb-Douglas. To resolve the question of which function is a better representation of data under OLS, we perform an F-test on the implied restrictions on Cobb-Douglas that the coefficients of $(\ln \text{area})^2$, $(\ln \text{fert})^2$, and (ln area*ln fert) are all zero. The calculated F-statistic = 47.7 which is highly significant at 1 percent level. Thus, we accept the translog function to reflect reality better.

Moving on to the stochastic frontier version of the translogarithmic model, and using the same inputs, we obtain the following results:

$$
\begin{aligned}
\widehat{\ln crpval_i} = {} & \underset{(0.184)}{7.92^{***}} + \underset{(0.086)}{0.657}\ln area_i^{***} + \underset{(0.036)}{0.192}\ln fert_i^{***} + \underset{(0.011)}{.018}(\ln area)^{2*} \\
& + \underset{(0.004)}{0.0452}(\ln fert)^{2***} - \underset{(0.009)}{0.0527}(\ln area * \ln fert)^{***}
\end{aligned} \tag{8}
$$

$$\log L = -2984.6, \quad \sigma_v = 0.581, \quad \sigma_u = 0.894, \quad \sigma^2 = \sigma_u^2 + \sigma_v^2 = 1.137,$$

$$\lambda = \sigma_u / \sigma_v = 1.537, \quad \pi = \sigma_u^2 / (\sigma_u^2 + \sigma_v^2) = 0.703, \quad N = 2535$$

where the parentheses below the coefficients indicate the standard errors underlying z-statistics. Unlike with OLS where it was insignificant, the coefficient of $(\ln \text{area})^2$ now passes the test at 10 percent, while other coefficients stay highly significant. The elasticity values for output with respect to inputs in the frontier estimation (0.654 and 0.155) undergo no substantial changes from their levels in the OLS regression.

The frontier estimation provides several other useful statistics. The estimate of π shown conceptually in equation (4) yields the proportion of idiosyncratic shocks specific to farmers to the total shocks that include shocks beyond the farmers' control. We find this statistic for Nepali farmers to be equal to 0.703, that is, 70 percent of the total variance in *u* and *v* is attributable to *u* alone. Moreover, the ratio of the standard deviations (σ_u/σ_v) equals 1.54 which implies a substantial range of inefficiency among farmers. The average inefficiency, given the prevailing modes of production, equals 38 percent of the maximum output based on the frontier estimates, since the mean efficiency is 62 percent of the maximum, with a standard deviation of 15.2 percent.

In the estimation of the farmer-specific efficiency levels, we use various characteristics of the households in the sample. Our main results appear in equation (9):

$$
\begin{aligned}
TE_i = \underset{(0.015)}{0.5097} &+ \underset{(0.0002)}{0.0005} agehd^{**} + \underset{(0.0077)}{0.0202} sexhd^{***} - \underset{(0.0056)}{0.00004} eduhd - \underset{(0.0043)}{0.018} occuhd^{***} \\
&+ \underset{(0.0009)}{0.002} fertpest^{**} + \underset{(0.0063)}{0.0219} irridm^{***} + \underset{(0.008)}{0.0524} areamed^{***} + \underset{(0.0079)}{0.0448} arealrg^{***} \qquad (9) \\
\bar{R}^2 = 0.038, &\qquad F = 13.39, \qquad N = 2535
\end{aligned}
$$

where the variable suffix '*hd*' means the head of household, *fertpest* is the interaction between the use of fertilizers and pesticides, *irridum* is the irrigation dummy (0: no irrigation, 1: yes), *areamed* and *arealrg* are dummies for medium and large farm sizes respectively, where arealrg=1 if *cropped* area is greater than 3 hectares, and *areamed*=1 if the area is greater than 1 hectare and less than or equal to 3 hectares.

As expected, the interaction between fertilizers and pesticides raises efficiency. The inclusion of the pesticide use separately in addition to the interaction between these two factors does not make a substantial change in our results (full results available upon request). This variable comes out statistically significant but makes the interaction coefficient insignificant, and leaves all other coefficients and their standard errors virtually unchanged. Among other results shown in equation (9), efficiency rises with the age of the household. When we include the age squared to check if the effect of age is nonlinear, we do find a small (−0.00004) but significant negative coefficient (at one percent level). Thus, at sufficiently old age of the head of the household, efficiency begins to fall. Households headed by a male also have slightly higher efficiency but the education variable produces no effect on efficiency after we control for age, sex and occupation.

Furthermore, irrigation dummy is highly significant although its size is rather small. This, however, does not resolve the question of a differential impact of irrigation across farms of different soil quality, or whether irrigation is available in a few months or year long or its possible interaction with high-yielding seeds, fertilizers and pesticides. It is also possible to explore interaction of irrigation with institutional (including tenancy) arrangements in Nepali farming.

Finally, we do find a substantial improvement in efficiency for medium and large farmers compared to small farmers. On average, households with larger farms achieve about 10 percent greater efficiency than small farms of up to one hectare even after irrigation and other variables are controlled for as in equation (9). It is important to note, however, that large farms do not seem to gain any particular advantage over medium farms in the country.

Conclusion

We find that the translog production function represents the NLSS data on farming better even though output elasticity estimates from the Cobb Douglas function also come close. The stochastic frontier estimate yields the separation of the effects of household-specific shocks from random shocks that affect Nepali agriculture in general. The average level of efficiency in Nepal's crop production is about 62 percent of efficiency achieved by the best practice farms. A mix of household characteristics together with size of farms impinges on farm efficiency in Nepal.

A main limitation of our study comes from limited nature of our data set. In particular, a more thorough processing of data can determine the values of physical capital used in farming. A major problem was that we could not arrive at suitable numbers or values for oxen or other draught animals used in farming. The aggregative number of animals in the raw NLSS data included chickens and goats which were of no use in crop production.

Another caveat relates to our division of farms into small, medium and large. However, the criteria set at 1 and 3 hectares, while generally sensible in the context of Nepal, are still arbitrary, and using data to devise a different scheme can change our results, at least to some degree.

We pursue further work along several other lines as well. One is to see differences in technical efficiency among the three topographical and five development regions of the country. Another is to make a greater use of education and health data within the family, rather than be restricted to one bit of information on education, namely the education of the household head. Further, the access to and use of extension services can indicate the degree to which the extension policy has been effective.

References

Battese, G. E. and Coelli, T. J. (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function and Panel Data. *Empirical Economics*, 20, 325-332.

Central Bureau of Statistics (CBS). (2004) *Nepal Living Standards Survey 2003/04* Statistical Report (in two Volumes). Kathmandu: CBS.

Jondrow, J., C. A. K. Lovell, I. S. Materov and P. Schmidt (1982) “On the estimation of technical inefficiency in the stochastic frontier production function model”, *Journal of Econometrics*, Vol. 19, pp. 233-238.