A Note on Production Function Estimation with Selectivity and Risk Considerations

Phoebe Koundouri and Céline Nauges

Dr. Phoebe Koundouri is a senior lecturer in Economics at the School of Business, Department of Economics, University of Reading (UK) and an Honorary Research Fellow at the Department of Economics/CSERGE, University College London (UK). Dr. Céline Nauges is a researcher at the French National Institute for Research in Agriculture (INRA), Toulouse (France).

The authors are grateful to Ben Groom for his work on preparing the data and to Alban Thomas, the journal editor and two anonymous referees for their insight and comments. The usual disclaimer applies.
A Note on Production Function Estimation with Selectivity and Risk Considerations

Abstract

In the estimation of production functions, ignoring risk considerations can cause inefficient estimates, whilst biased parameter estimates arise in the presence of sample selection. In the presence of uncertainty and selection bias, the latter introduced by the endogeneity of qualitative characteristics of inputs in crop choice, we show that correcting for risk considerations (à la Just and Pope 1978, 1979) but not selection bias, can produce incorrect inferences in terms of risk behavior. The arguments raised in this paper have estimation and policy implications for stochastic production analysis applied to all goods whose qualitative characteristics can affect sample selection.

Keywords: crop choice, production risk, sample selection.
1. Introduction

Just and Pope (JP from now on) (1977, 1978) developed general models for handling production risk econometrically. Their approach has been quite popular among agricultural economists, and is still prominently used in the recent literature (Tveterås; Kumbhakar and Tsionas, among others). JP’s basic idea was to write the production function as the sum of two components, one relating to the output level, and one relating to the variability of output. This specification allows the econometrician to differentiate the impact of input on output and risk, and has sufficient flexibility to accommodate both positive and negative marginal risks with respect to inputs. In addition, JP show that ignoring risk in the production function can lead to wrong inferences on the technology coefficients and in particular can produce standard errors that are misleading in that they indicate much greater precision in estimation than is, in fact, obtained. In this note, we extend JP’s result by showing, through a simple empirical illustration using a Just-Pope production function, that the parameters of the risk function can be inconsistent if we fail to control for selectivity in crop choice, i.e. specific characteristics of fixed and quasi-fixed inputs, observed and unobserved (e.g. quality of groundwater, type of soil, distance to town), could affect both the choice of the crop to be cultivated and the production function.

Using a simple framework in which farmers are assumed to be risk neutral and in which risk only enters the estimation of the crop-specific production functions, we illustrate biases that might affect the risk function parameters if one does not correct for selectivity in crop choice. Although in real production processes farmers are likely to be risk averse and risk is likely to play a role in the crop-choice decision process, we do not explicitly account for this due to data limitations. We briefly discuss, however, the ‘correct’ theoretical approach and the implied data requirements.
Selectivity, which was not an issue in JP’s articles as they used experimental data, might be an issue in empirical applications using economic data. Selectivity has been acknowledged and taken into account in numerous agricultural models, but always in a risk-free environment. Among others, Moore, Gollehon, and Carey consider selectivity in modelling crop choice decision in their analysis of multi-output production functions. Heshmati shows that not correcting for selectivity (due to aggregation/truncation from the original sample) can bias the estimation of relevant parameters such as input elasticities and returns to scale. Selectivity has also been commonly considered in the agricultural literature on adoption (Dinar and Zilberman, Khanna).

The paper is organized as follows. In section 2 we model the simultaneity between the production function and sample selection introduced by crop choice, in the context of a single-output model of farmer behavior with production risk. In section 3 we apply this model to a Just-Pope production function with selectivity corrections using the two-step Heckman’s method, on a farm-level sample from Cyprus. Our results suggest that failing to correct for sample selection produces biased parameter estimates of the risk function. These results have important implications for agricultural policy, as they show that when selection bias is present, the impact of a policy instrument, both in terms of (foregone) expected profit from production and in terms of the change in revenue as a consequence of hedging against modified production risk, could be misleading.

2. Theoretical description

We ignore production risk for the moment and we consider a representative producer facing a choice set of \( L \) crops suitable for production. We focus on the case where the outcome of the
producer’s optimization problem is a ‘corner solution’, representing the decision to cultivate one particular crop. To keep the analysis simple, we assume that the farmer is risk-neutral and hence maximises the mathematical expectation of profit. The farmer decides which crop to cultivate among a set of possible crops, by comparing his expected profit, $\pi^*$, for different cultivation choices. Using the dummy variable $D_l = 1$ when the $l$th crop is selected by the farmer and $D_l = 0$ otherwise, we can write

$$D_l = 1 \text{ if } \pi_l^* > 0 \text{ and } \pi_l^* = \max(\pi_1^*, \ldots, \pi_L^*)$$

and $D_l = 0$ otherwise,

where $\pi_l^*$ represents the expected profit when crop $l$ is chosen. $\pi_l^*$ is unobservable when crop $l$ is not grown (our data contains a single cross-section of farmers, hence we observe each farmer in a unique situation, i.e. growing a single type of crop), hence we model the decision of the farmer to grow crop $l$ by using a discrete choice approach: $D_l = 1[g(z, \lambda_l) + \nu_l > 0]$

where $g(.)$ is an unknown function of variables $z$ and unknown parameters $\lambda_l$. We assume that differences in terms of expected profitability will be driven by variables, $z$, such as the environment of the farm (quality of accessible groundwater supplies, distance from town, rainfall, total irrigated area, etc.), but also characteristics of the farmer such as his experience in farming. As some of these characteristics may be unobservable, we specify an additive error term, $\nu_l$, assumed of mean 0. Once crop $l$ has been selected, the production function corresponding to this particular crop can be defined by

$$y_l = f(x, \beta_l) + u_l \text{ if } D_l = 1$$

where $y_l$ is a measure of output level for crop $l$ and $x$ is a vector gathering inputs as well as other shifters of the production function ($x$ may have some elements in common with $z$). $u_l$ is the usual econometric error term, assumed of mean 0.
With $D_l$ and $z$ observed for a random sample, but $y_l$ observed only when $D_l = 1$, the output variable in this equation is incidentally truncated from below, on a non-positive net profit from cultivating a particular crop (i.e. $\pi^*_l > 0$). Assuming that $v_l$ and $u_l$ have a bivariate normal distribution with zero means and correlation $\rho_l$

$$E[u_l|D_l=1] = \sigma_l \frac{\phi\left(g(z, \hat{\lambda}_l)\right)}{1 - \Phi\left(g(z, \hat{\lambda}_l)\right)}$$

where $E[\cdot]$ denotes expectations, $\phi(.)$ and $\Phi(.)$ represent the standard normal probability density and cumulative distribution functions respectively; $\sigma_l = \text{cov}(v_l, u_l)$ captures simultaneity in the participation and production function equations.

$M_l = \phi\left(g(z, \hat{\lambda}_l)\right)\left[1 - \Phi\left(g(z, \hat{\lambda}_l)\right)\right]$ is the so-called Mill ratio used in the two-step Heckman correction for selection bias. Therefore, for $D_l = 1$, we may write (2) as

$$y_l = f(x, \beta_l) + \sigma_l M_l + w_l$$

where $w_l$ is an error term assumed of mean 0.

Risk deriving from yield uncertainty is now incorporated in the production function as in JP (1979)

$$y_l = f(x, \beta_l) + \sigma_l M_l + w_l$$

with

$$w_l = h(x, \xi_l) \eta_l.$$  

$f(.)$ is the deterministic part of the production function and is called the mean function. $h(.)$ is the variance or risk function, which captures the effects of each input on the risk of production (as measured by the variance of output). The only random component in the model is $\eta_l$, which is assumed i.i.d. $N(0,1)$. The production disturbance, $\eta_l$, represents factors such
as weather, unpredictable variations in machine or labor performance. Along the lines of Just and Pope, we assume that $\eta$ is not known to the farmer at the time input decisions are made.

We choose a very simplified framework where we assume that production risk is the only source of risk (in particular we assume that there is no uncertainty on future input and output prices) and that the farmer is risk neutral. The assumption of farmer’s risk aversion would make the problem very complicated as one should model the discrete crop choice decision based on comparisons of expected utility levels. To implement such an approach, one needs to identify the parameters of risk preferences. This requires data on repeated cross-sections and the optimal input choices for alternative crops to be observed. The lack of data on repeated cross-sections, prevents us from making the analysis under the assumption of farmers’ risk aversion.

3. Empirical Analysis

3.1. Dataset description

A cross-section of 239 farms located in the agricultural region of Kiti in Cyprus was surveyed and farmers were asked to provide accurate information regarding production activities on representative parcels of their land. Information was provided regarding expenditures upon fixed and variable inputs used in the production of final outputs. Examples include pesticides, fertilizers, labor and water. Input prices, which were not observed at the farm level, are assumed to be homogeneous across farms, which is a reasonable assumption for a small agricultural region with many competitive farms, such as Kiti. Output quantities and prices are crop-specific. Variability in input and output levels across farms is guaranteed by the variability in environmental (climate, soil type, etc.) and socio-economic (education, age, etc.) conditions. Qualitative data was also provided, e.g. information regarding farm ownership,
family characteristics and access to water resources. The survey also included climatic data in the form of rainfall measurements.

The wide variety of crop types represented in the sample, necessitated the grouping of crops into broad categories in order to overcome the sparseness of individual crop observations. We choose to group crops into three categories, namely: vegetables (95 observations), cereals (89 observations), and citrus (55 observations). The reason we group outputs into these three categories is twofold: first, farms specialize either in vegetable, cereal, or citrus cultivation, and second, we only have information on total expenditure and total revenues of farms (i.e. we do not have such information for each specific crop). In what follows, we will focus on vegetables and cereals groups only. These two groups of crops have a number of characteristics that makes their cultivation structurally different. In particular, the production of vegetables is labour intensive (whereas cereals’ production is not) and vegetables are by far the most profitable crops cultivated in Cyprus, as they are competitive in the international arena (see Table 1 for relevant descriptive statistics).

3.2. Model specification

The production function is assumed to be of the JP form:

\[
y = f(x, \beta) + h(x, \xi)\eta
\]

where \(y\) measures output, \(x\) is the vector gathering variable inputs and extra production shifters. We assume \(E(\eta) = 0\) and \(V(\eta) = 1\). \(f(.)\) and \(h(.)\) are the mean function and the risk function, respectively. This specification allows the input vector to influence both the mean output and the variance of output, without any requirement on the sign of these effects (i.e. inputs can be risk-increasing or risk-decreasing). We choose a linear quadratic form for the
mean output function $f(.)$ for two reasons: first, because it is consistent with JP postulates (there is an additive interaction between the mean and variance output functions) and second, because it is flexible in the sense of a second-order approximation of any unknown mean-output function (Kumbhakar and Tveterås). The mean output function $f$ for the representative farm is written:

$$f(x, o) = \beta_0 + \sum_j \beta_j x_j + \sum_j \beta_{2j} x_j^2 + \sum_j \sum_{k \neq j} \beta_{jk} x_j x_k + o' \gamma,$$

where the vector of inputs $x$ gathers fertilisers and manure, pesticides, labour, and irrigation water. The vector $o$ denotes a vector of extra production shifters including total rainfall, total irrigated area (that we use as a proxy for soil type), total present value of investment in machinery, distances to the nearest town and coast, and years of experience in farming. To avoid multicollinearity, these variables are assumed to enter the mean production function linearly.

The variance or risk function $h(.)$ is modelled as a Cobb-Douglas, along the lines of JP (1978, 1979), and Kumbhakar and Tveterås:

$$h(x) = \xi_0 \prod_j x_j^{\xi_j},$$

where the $\xi$'s are parameters to be estimated.

3.3. Estimation procedure

Our purpose is to estimate the risk function for the two groups of farmers, assuming that their decision to grow either vegetables or cereals may also depend on observed and unobserved characteristics of the farm. If these characteristics also enter the production function, then parameter estimates of the risk function may be biased. Note that the fact that farms specialize either in vegetable or cereal cultivations, allows us to use a single (and not a multi-output) production function.
The estimation method consists of the three following steps:

**Step 1:** We estimate two selection equations, which predict the following two probabilities, respectively: (a) the probability of a farmer growing vegetables against the probability of growing any other crop (i.e. cereals or citrus); (b) the probability of a farmer growing cereals against the probability of growing any other crop (i.e. vegetables or citrus).

For both of these selection equations, \( g(.) \) is assumed linear in the parameters \( \lambda \), and \( v \) is assumed normally distributed. Referring back to the notation we used in equation (1), we estimate the probability that \( D_l = 1 \), where \( l \) indicates either vegetables or cereals, using a Probit model. We assume that the choice of crop is driven by the environment of the farm (rainfall, distances to nearest town and coast), characteristics of the farmer such as his experience in farming, and total irrigated area (here again used as a proxy for soil type). The Mill’s ratio for each of the crop-specific production functions is computed from the estimated parameters, \( \hat{\lambda} \), of the respective crop-specific Probit model.

**Step 2:** Ignoring the presence of the risk-function \( h(.) \) in the production function model (see equation 4), we estimate the crop-specific production functions by Generalized Method of Moments (GMM) replacing the unknown Mill’s ratios by their estimated value. As our data-set is a single cross-section containing only input expenditures and no separate information for price and input quantity, we estimate the production functions assuming that the prices of inputs are homogenous. We choose to work on crop yield per unit of surface and not on total production in order to eliminate size effects.\(^8\)

The production function is estimated separately from the input demand equations as it can be easily shown that derived input demands do not depend on the production shock \( \eta \) when one assumes that the production shock is not known to the farmer at the time input decisions are made (Zellner, Kmenta and Drèze).\(^9\)
GMM, which is robust to heteroskedasticity, produces unbiased and efficient estimates of the set of parameters (\( \beta, \gamma, \sigma \)).

**Step 3:** Estimate the residuals

\[
\hat{w} = y - \left( \hat{\beta}_0 + \sum_j \hat{\beta}_j x_j + \sum_j \hat{\beta}_2 j x_j^2 + \sum_j \sum_{k(k \neq j)} \hat{\beta}_{jk} x_j x_k + \theta \hat{\gamma} + \hat{\sigma} \hat{M} \right)
\]

which are then used to estimate the risk function which is written as in JP (1979)

\[
\ln |\hat{w}| = \xi_0 + \sum_j \xi_j \ln (x_j) + \ln \eta, \text{ see equations (6) and (9).}
\]

GMM provides efficient and unbiased \( \xi \) coefficients.

### 3.4. Estimation results

Probit analyses for vegetables and cereals are reported in Tables 2 and 3, respectively. Results indicate that the choice of cultivating either vegetables or cereals is influenced by the qualitative characteristics of the fixed and quasi-fixed inputs of production, as modelled in section 2.

[Tables 2 and 3 here]

The probability of cultivating vegetables [resp. cereals] is positively [resp. negatively] affected by the proportion of the parcel of land being irrigated, as vegetables cultivation requires more water than cereals (see Table 1). Distance from the coast [resp. from the town] decreases [resp. increases] the probability of vegetables cultivation. We find the opposite effects in the case of cereals. Finally, we find evidence that more experienced farmers are more likely to grow cereals and less likely to grow vegetables. From these estimates, we compute, for each sub-group (vegetables producers and cereals producers) the Mill’s ratio that we incorporate in each production function. In addition to the Mill’s ratio, we include as determinants the following variables: variable inputs (pesticides, labor, water, fertilisers),
investment in machinery, rainfall, total irrigated area, distance to nearest town, distance to nearest cost, and years of experience in farming. The production function is estimated separately for both groups of farmers. In each case the estimated model exhibits a satisfactorily high adjusted-$R^2$ for cross-section data, 0.80 and 0.83 for the vegetables and cereals group, respectively. The Mill’s ratio is found significant (at the 1% level of significance) for the sub-group of vegetables growers, whereas it is significant at the 20% level of significance only for the sub-group of cereals producers. Estimation results of the mean production function in the two sub-groups of farmers are shown in Table A1 in Appendix.

Tables 4 and 5 contain the parameters of the risk function $h(.)$ estimated with and without selectivity correction for all variable inputs, for both the vegetables and cereals groups. Note that all standard errors have been bootstrapped using 500 replications. This is a finite sample correction on the estimated errors.

[Tables 4 and 5 here]

The aim of this exercise is to investigate how risk analysis is affected by ignoring selectivity bias in the estimation of the production function. From Tables 4 and 5, failing to correct for endogeneity in crop choice is found to bias parameter estimates. For the vegetables group, even though the sign of the effects remains the same, the contribution of each input to the variance is found to be different depending on whether or not we correct for selectivity. More precisely, although labor is found to be risk-increasing irrespective of whether or not we control for selectivity, the impact of pesticides and fertilisers is different depending on whether or not we correct for selectivity. Fertilisers are proved not to affect production risk significantly when selectivity is not taken into account, but is found to have a positive and significant effect on risk when we correct for selectivity bias. In the case of pesticides, this
input is found to be risk-increasing only when selectivity is taken into account. The value of
the estimated coefficients, which can be interpreted as output variance elasticities (Tveterås,
2000), also varies between selectivity corrected and non-selectivity corrected models.
Variance elasticity with respect to labor is estimated at –0.36 when there is a correction for
selectivity while it is estimated at –0.30 when there is not. The bias in the measurement of
variance elasticity is larger in the case of pesticides: the unbiased elasticity is found equal to
0.27, while it is estimated at 0.19 when there is no correction for selectivity. For the case of
cereals, we observe some differences between estimates with and without selectivity
correction, even if the Mill’s ratio was almost non-significant in the production model for
cereals producers. Labor and water are found to be risk-decreasing inputs in both models
(high level of significance) however the magnitude of their effects varies from one model to
the other. Not also that, even though it is not significant, the estimated impact of pesticides is
found positive in one case (the case with selectivity correction) and negative in the other.

We have to remain cautious about the estimates of variance elasticities as our model has been
estimated under the assumption that farmers are risk neutral. However we argue that this
simple illustration shows that failing to correct for sample selection in the estimation of a
stochastic production function leads to wrong inferences as far as the risk function is
concerned. Such bias has important policy implications as it can produce misleading
inferences on input-specific risk effects. For example, consider a policy maker who is
contemplating the introduction of a conservation policy aiming to reduce the use of fertilizers,
in order to restrict the adverse effects of such use on the quality of soil and groundwater
resources and adhere to the relevant European Union environmental directive. If this policy
maker decides to introduce a quota on the use of these inputs, then he should first investigate
how this policy instrument will affect agricultural revenue not only in terms of (foregone)
expected profit from production alone, but also in terms of the change in revenue as the consequence of hedging against a modified production risk (the latter originating from the need of the farmer to modify his insurance behavior). In such a case, agricultural subsidy schemes should be designated simultaneously with the environmental quota. If, however, selection bias exists in policy simulations, then produced policies will be inefficient not only with respect to technical issues but also with respect to risk hedging.

Conclusion

This paper investigates selectivity bias introduced by crop-choice in agricultural production under risk. To avoid estimation bias from ignoring sample selection biases, we apply Heckman's correction to a Just-Pope production function. The empirical analysis runs under the assumption of farmer’s risk neutrality and suggests that failing to correct for sample selection, biases estimation of input-specific marginal risk. The arguments raised in this paper have implications for stochastic production analysis applied to all goods whose qualitative characteristics can affect sample selection.
Endnotes

1. For details on the construction of the questionnaire and collection of the data set, see Koundouri.

2. Data on the quantities of water used in crop production were sparse and often inconsistent. In response, information regarding water requirements for the specified crops were gathered from the Ministry of Agriculture (Agricultural Research Institute) and were used to calculate theoretical water demands for the farms based on the areas of land devoted to particular crops. This information is used when farm-specific data on irrigation water is missing. Although one of the questions in the questionnaire concerned water use and water costs, the responses for these particular questions were sparse and did not reveal the marginal costs faced by individual farmers. In response to this we have constructed a tariff for groundwater pumping costs based on hydrological information obtained from the Ministry of Agriculture in Nicosia.

3. We do not present the results for the sub-group of citrus producers because of identification problems and lack of fit of the production function.

4. Note that because some information on expenditure are missing for 3 vegetables producers we report, in Table 1, descriptive statistics based on 92 producers only.

5. From this point onwards we suppress subscript \( l \) for the crop.

6. This would not have been the case for a translog specification, which assumes a multiplicative interaction between the mean and variance functions.

7. Irrigated area is determined by soil type, and for this reason it can be assumed exogenous in the model.

8. We also estimated a production function incorporating absolute levels of production and inputs, along with the overall cultivated surface, in order to overcome the assumption of
constant returns to scale. However, the surface variable turned out to be the only significant variable in the estimation due to the high collinearity present in the data.

9. There has been a large debate on the issue of input endogeneity in the estimation of production function with risk considerations following the article by Love and Buccola in 1991. These authors argue that when input choice is influenced by perceived production risk, the latter being defined in terms of the production error, correlation between inputs and the error term is likely to be present in production models. Shankar and Nelson proved that inconsistency was not an issue in the version of the JP production model in which both the mean and the variance functions have the Cobb-Douglas form.
References


<table>
<thead>
<tr>
<th></th>
<th>Vegetables</th>
<th></th>
<th>Cereals</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>St Dev.</td>
<td>Mean</td>
<td>St Dev.</td>
</tr>
<tr>
<td>surface allocated (A)(^a) (ha)</td>
<td>3.23</td>
<td>6.76</td>
<td>3.13</td>
<td>4.60</td>
</tr>
<tr>
<td>gross revenue/ha (CYP/a)(^b)</td>
<td>2,786</td>
<td>3,764</td>
<td>745</td>
<td>1,331</td>
</tr>
<tr>
<td>fertilisers (F)(^c) expend. (CYP/ha/a)</td>
<td>243.13</td>
<td>444.26</td>
<td>119.73</td>
<td>325.13</td>
</tr>
<tr>
<td>pesticides (P) expend. (CYP/ha/a)</td>
<td>152.18</td>
<td>419.03</td>
<td>51.08</td>
<td>151.09</td>
</tr>
<tr>
<td>labor (L)(^d) expend. (CYP/ha/a)</td>
<td>229.91</td>
<td>714.42</td>
<td>105.36</td>
<td>211.99</td>
</tr>
<tr>
<td>water (W) expend. (CYP/ha/a)</td>
<td>117.92</td>
<td>600.52</td>
<td>50.03</td>
<td>194.42</td>
</tr>
</tbody>
</table>

Number of observations 92 89

Notes:
- a. includes irrigated and non-irrigated areas.
- b. CYP: Cyprus pound (1 CYP was around 1.5 US Dollar at the time of the survey). “a” is for annum.
- c. including manure.
- d. casual work in production.
Table 2: Estimation of the probability of growing vegetables – Probit analysis

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>Std Err.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>1.2990</td>
<td>0.6135</td>
<td>0.034</td>
</tr>
<tr>
<td>Rainfall in 1998 (mm)</td>
<td>-0.0323</td>
<td>0.0221</td>
<td>0.145</td>
</tr>
<tr>
<td>Total irrigated area</td>
<td>0.0055</td>
<td>0.0049</td>
<td>0.268</td>
</tr>
<tr>
<td>Distance to nearest town (km)</td>
<td>0.0348</td>
<td>0.0158</td>
<td>0.028</td>
</tr>
<tr>
<td>Distance to nearest coast (km)</td>
<td>-0.0772</td>
<td>0.0378</td>
<td>0.041</td>
</tr>
<tr>
<td>Years of experience in farming</td>
<td>-0.0290</td>
<td>0.0085</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Observations 239

Likelihood Ratio test statistic 20.66
(p-value) (0.0009)
Percentage of correct predictions 63.60
Table 3: Estimation of the probability of growing cereals – Probit analysis

<table>
<thead>
<tr>
<th></th>
<th>Coef.</th>
<th>Std Err.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.4641</td>
<td>0.6302</td>
<td>0.020</td>
</tr>
<tr>
<td>Rainfall in 1998 (mm)</td>
<td>0.0194</td>
<td>0.0225</td>
<td>0.388</td>
</tr>
<tr>
<td>Total irrigated area</td>
<td>-0.0125</td>
<td>0.0060</td>
<td>0.035</td>
</tr>
<tr>
<td>Distance to nearest town (km)</td>
<td>-0.0710</td>
<td>0.0180</td>
<td>0.000</td>
</tr>
<tr>
<td>Distance to nearest coast (km)</td>
<td>0.1029</td>
<td>0.0385</td>
<td>0.007</td>
</tr>
<tr>
<td>Years of experience in farming</td>
<td>0.0371</td>
<td>0.0093</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Observations 239

Likelihood Ratio test statistic 41.27
(p-value) (0.0000)

Percentage of correct predictions 64.44
Table 4: Estimation of the risk function for the group of vegetables growers

<table>
<thead>
<tr>
<th></th>
<th>With selectivity correction</th>
<th>Without selectivity corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fertilisers</td>
<td>0.449</td>
<td>0.257</td>
</tr>
<tr>
<td>Pesticides</td>
<td>0.186</td>
<td>0.136</td>
</tr>
<tr>
<td>Labor</td>
<td>-0.361</td>
<td>0.078</td>
</tr>
<tr>
<td>Water</td>
<td>0.055</td>
<td>0.170</td>
</tr>
</tbody>
</table>

a: bootstrapped standard errors computed using 500 replications.
Table 5: Estimation of the risk function for the group of cereals growers

<table>
<thead>
<tr>
<th></th>
<th>With selectivity correction</th>
<th>Without selectivity corrections</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std Err. (^a)</td>
</tr>
<tr>
<td>Fertilisers</td>
<td>0.516</td>
<td>0.565</td>
</tr>
<tr>
<td>Pesticides</td>
<td>0.048</td>
<td>0.279</td>
</tr>
<tr>
<td>Labor</td>
<td>-0.365</td>
<td>0.086</td>
</tr>
<tr>
<td>Water</td>
<td>-1.031</td>
<td>0.531</td>
</tr>
</tbody>
</table>

\(^a\): bootstrapped standard errors computed using 500 replications.
## Appendix

Table A1: Estimation of the mean production function (including Mill ratio)
on both groups of farmers

<table>
<thead>
<tr>
<th></th>
<th>Vegetables group</th>
<th>Cereals group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
<td>Std. Err.</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.595</td>
<td>0.889</td>
</tr>
<tr>
<td>Fertilisers (F)(^a)</td>
<td>0.003</td>
<td>0.200</td>
</tr>
<tr>
<td>Pesticides (P)</td>
<td>0.425</td>
<td>0.131</td>
</tr>
<tr>
<td>Labour (L)</td>
<td>-0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>Water (W)</td>
<td>1.357</td>
<td>0.288</td>
</tr>
<tr>
<td>FxF</td>
<td>0.121</td>
<td>0.046</td>
</tr>
<tr>
<td>PxP</td>
<td>0.000</td>
<td>0.035</td>
</tr>
<tr>
<td>LxL</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>WxW</td>
<td>-0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>FxP</td>
<td>-0.060</td>
<td>0.089</td>
</tr>
<tr>
<td>FxL</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>FxW</td>
<td>-0.011</td>
<td>0.039</td>
</tr>
<tr>
<td>PxL</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>PxW</td>
<td>-0.657</td>
<td>0.137</td>
</tr>
<tr>
<td>LxW</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Investment in machinery</td>
<td>-0.030</td>
<td>0.068</td>
</tr>
<tr>
<td>Total irrigated area</td>
<td>-0.614</td>
<td>0.192</td>
</tr>
<tr>
<td>Total rainfall in 1998</td>
<td>11.361</td>
<td>3.158</td>
</tr>
<tr>
<td>Distance to nearest town</td>
<td>-5.379</td>
<td>1.560</td>
</tr>
<tr>
<td>Distance to nearest coast</td>
<td>3.048</td>
<td>0.954</td>
</tr>
<tr>
<td>Years of experience</td>
<td>9.463</td>
<td>2.782</td>
</tr>
<tr>
<td>Mill ratio</td>
<td>-16.619</td>
<td>4.978</td>
</tr>
</tbody>
</table>

Number of observations 92 89  
Adjusted R-squared 0.80 0.83

\(^a\) all variables have been mean-scaled.