ESTIMATING GENDER DIFFERENCES IN AGRICULTURAL PRODUCTIVITY:

BIASES DUE TO OMISSION OF GENDER-INFLUENCED VARIABLES

AND ENDOGENEITY OF REGRESSORS

by

Nina Lilja, Thomas F. Randolph and Abrahmane Diallo*

Selected paper submitted to

the American Agricultural Economics Association Annual meeting in

Salt Lake City, Utah

August 2-5, 1998

*Dr. Lilja is a post-doctoral fellow at the West Africa Rice Development Association (WARDA), 01 BP 2551, Bouaké 01, Côte d'Ivoire, E (N.LILJA@cgent.com). Dr. Randolph is a policy economist in Côte d'Ivoire. Mr. Diallo is a consultant at the African Consultants in Social Sciences for Development, BP 16123 Dakar-Fann, Dakar, Sénégal. All correspondence regarding this paper should be sent to Dr. Lilja. Authorship is shared jointly.

ESTIMATING GENDER DIFFERENCES IN AGRICULTURAL PRODUCTIVITY: BIASES DUE TO OMISSION OF GENDER-INFLUENCED VARIABLES AND ENDOGENEITY OF REGRESSORS

Introduction

Most empirical estimates of male-female agricultural productivity differences show that, controlling for input levels and human capital, male and female farmers are equally efficient farm managers. In a recent article, Quisumbing (1996) has noted that many empirical gender productivity studies are methodologically flawed. Specifically, she points out that greater attention needs to be given to appropriate estimation methods, choice of "gender variables" and endogenous input choice. She notes, that the inadequate specification of individual farmer characteristics in empirical studies has often resulted in confusion between "sex" and "gender" as a source of productivity differences. ("Sex differences" are due to biological differences between men and women whereas "gender differences" are influenced by ideological, religious, ethnic, economic and cultural determinants which contribute to, for example, differential access to productive resources and information between men and women farmers.) Improved representation of gender factors in the analysis of farm-level productivity is essential if productivity differentials associated with gender are to be properly identified and the sources of those differentials correctly diagnosed. Such information is essential for the formulation of appropriate policies to enhance women's agricultural productivity.

The purpose of this study is to isolate managerial differences between men and women in an agricultural setting where male and female farmers cultivate a similar crop and have access to similar technology, and to assess the degree of bias introduced by omission of gender-influenced variables and the endogeneity of input levels. Diallo's work in Senegal (1998) provides a rich set of anthropological and agro-economic data regarding farmer characteristics and knowledge, resources, and managerial ability as reflected by timing of farming activities. Using plot-level observations for a pooled sample of male and female-managed irrigated rice plots, our paper estimates three types of models. In each model, an equation for rice yield per hectare of cultivated land as the dependent variable is estimated using plot-level observations, disaggregated by the sex of the plot manager. Information generated by comparison of these three models allows us to determine the degree of bias due to omission of gender-influenced variables or simultaneous-equation bias. This study also allows us to assess the importance or necessity of collecting time-consuming and financially costly gender-diasaggregated anthropological data. The results also give insight to improving research methods for isolating the gender impact on productivity.

Empirical Models

Our first model, the *basic model*, is designed to include the variables commonly used in male-female agricultural productivity studies that are generated from the data collected in standard economic surveys, and estimates the yield impact of inputs used (labor, fertilizer, irrigation, pest control) and individual farmer attributes (gender, education). Although this basic model controls for individual characteristics such as education and physical assets, the model may still overestimate productivity differences associated with sex of the farm-operator because gender dummy variable captures both "sex" and "gender" determined information.

Our second model, the *extended model*, is designed to deal with the model misspecification problem due to commonly omitted variables in female-male productivity estimations. It is often argued that women's domestic obligations as well as labor obligations to family fields often prevents women allocating adequate labor time to her own fields. Our *extended model* includes "gender-influenced variables" such as timing of farm operations (planting and weeding).

Because some of the regressors (input levels) are likely to be endogenously determined, they will be correlated with the error term and the Ordinary Least Squares (OLS) estimates will be inconsistent. We will discuss a third model, the *endogenous input choice model*, and explain our attempt to use instrumental variable technique to deal with simultaneous-equations bias.

Data

Data used in this study comes from Diallo's (1998) survey of 58 male and 77 female irrigated rice producers in Senegal. The survey used multiple visits guided by a series of questionnaires to collect information on factor and input use during the 1994 rainy season in a sample of irrigated rice plots, as well as other information about household characteristics and the respondents' other economic activities. The plots were located in six different village irrigated schemes in the Upper Senegal River Valley (Senegal), representing two different irrigation designs: (i) a first generation built during the 1970s on sandier soils; and (ii) a second, improved generation built on clay soils. The average rice area cultivated by women farmers was 0.12 ha compared to an average of 0.40 ha for men operators. The production system is semi-intensive; nearly all the farmers (98%) in the sample applied fertilizer, but only 10% of the farmers in the sample applied herbicide.

Input use and expenditures

Table 1 shows sample averages for input use and expenditures. Although the total labor use on fields operated by men and women is almost the same, men farmers hired more labor than women farmers. Assuming that women and women can hire labor at same wage rate, the data implies that the men hired twice as much total agricultural labor than women. Men spent 41% more on hired labor for weeding, and 21% more on harvesting labor than women, but women spent relatively more on hired labor for land preparation (20%) than men. On average, women's fields were irrigated one time less than men. Men spent 7% more on fertilizers on their fields than women farm operators, and applied 4% more (7 kg) more fertilizer on their fields than women. Men's fields were guarded against bird damage on average 5 days more than women's fields.

Table 1: Input use and expenditures in irrigated rice production, Senegal

Input/Expenditure category	Sample Average		% difference between women and men
	Women	Men	
Sample size	77	58	
Inputs:			
Labor (hours per ha)	312	303	3
Number of irrigation	5	6	-20
Number of days fields was guarded against birds	1	6	-500
Total fertilizer applied (kg/ha)	175	182	4
Expenditures:			
Land preparation labor cost (FCFA/ha)	22,555	18,134	20
Weeding labor cost (FCFA/ha)	12,731	17,916	-41
Harvest labor cost (FCFA/kg)	16,821	20,334	-21
Total fertilizer expenditure (FCFA/ha)	49,037	52,476	-7
% of total output produced for consumption	80	79	1
% of total output produced disbursed as gifts	9	4	5

Timing of agricultural activities

Table 2 shows the differences between male and female farm operators' ability to complete various activities before an internally defined reference date. The reference period is identified as the period when the middle tercile of the sample performed the operation. The most noticeable difference in timing of the farm activities is that 24% of the men's fields are weeded before the reference period, compared to only 4% of women's fields. Also, 35% of men's fields are harvested before the reference period versus 26% of women's fields.

Table 2: Share of sample performing activities in irrigated rice production before the reference period

Activity	Women	Men
Sample size	77	58
Early planting	31%	26%
Early weeding	4%	24%
Early first fertilizer application	33%	28%
Early second fertilizer application	22%	24%
Early harvest	26%	35%

Yield Differences in Irrigated Rice Production

The sample average yield was 2,380 kg/ha. In our sample, the average per hectare yield from the fields operated by men rice farmers was 32% higher than that of fields operated by women farmers. The empirical data suggests that major constraints that women farm operators face (compared to male farm operators) in irrigated rice production are: lack of irrigation water, not enough weeding labor or poor timing of the weeding and harvest, and bird damage.

Model Results

The linear models we are estimating are not production functions but rather reduced form equations which provide a test of impact of allocation of inputs, farmer characteristics and timing of farming activities across the plots controlled by men and women farmers. The finding that there are gender differences in yield does not necessarily imply that women are less efficient farmers than men, but may reflect differences in access to inputs, as well as the intensity and the timing associated with the application of the inputs onto women and men's rice fields.

In the *basic model* we estimate an equation: $y = \alpha + \beta X + \gamma D + \varepsilon$, where y is a rice yield per hectare, X is a vector of characteristics of the plot, the farmer, and input use, and includes the total amount of labor per hectare allocated to the rice plot, number of irrigations during the season, number of days the field was guarded against birds, and a dummy variable for irrigation perimeter and for fertilizer application (whether there was more than one application). The number of fertilizer applications is used rather than quantity applied since recent research has generally found this factor to be more important in explaining yield variation in irrigated rice systems in the lower Senegal River Valley (Dingkuhn, 1992). A dummy variable for the first generation irrigation schemes is included to capture the differences in soil type between the two generations of perimeters. The education level of the farmer is commonly used as an explanatory variable for farmer characteristics but due to the large male-bias in access to education in West Africa, the level of formal education does not appear to successfully capture the "farming knowledge" of the respondent. Instead, we use a score for general knowledge on agricultural practices as a proxy variable for this aspect of human capital. Respondents were asked a series of questions related to rice production techniques to derive a knowledge score. This variable is less gender-biased than education which is heavily influenced by restricted school attendance for females. The model also includes a dummy variable (D) for the gender of the plot operator, and ε is the error term.

The results of our *basic model* estimation are presented in Table 3 and show that the total amount of labor, number of irrigations and number of days the field was guarded against birds have a high and positive significant impact on yield. Some other explanatory variables were significant at 10% level indicating that more than one fertilizer application and gender of the farm operator have a positive significant impact on the yield. The gender coefficient suggests that there is a significant productivity difference due to unspecified gender-based and sex differences, other factors being equal, with female farmers obtaining an average 479 kg of paddy/ha less than male farmers, a 17% loss relative to the average yield for male farmers of 2,758 kg/ha. Farming knowledge and perimeter did not have a significant impact on the yield.

Table 3: OLS Results, Basic Model

Variable	Estimate	t-ratio
Labor (man hours/ha)	3.023**	2.993
Number of irrigations	129.53**	4.528
Number of days field was guarded against birds	35.494*	2.282
Split fertilizer application (1=more than 1 application)	504.05	1.870
Total farming knowledge score	0.021	1.455
Dummy Variable for Perimeter (1=1st perimeter)	-590.99	-1.509
Dummy Variable for the Gender of the Farmer (1=male)	479.27	1.887
Constant	333.24	0.974

Dependent variable= rice yield, kg per hectare. N=135 R^2=.37, R^2 adjusted=.34
**significant at 1% level, * significant at 5% level

In our *extended model*, we allow both slope and intercept to change between male and female farm operators.[1] Given the *basic model*, $y = \alpha + \beta X + \gamma D + \varepsilon$, we assume that the coefficient of gender dummy variable D is a function of other "gender-influenced" variables (Z), such that

[1]Because the model is expressed in a single equation, the variance of the error term is assumed to be same for male and female farm operators

$\gamma = \phi + \delta Z$. Substituting for γ, the model becomes: $y = \alpha + \beta X + \phi D + \delta ZD + \varepsilon$. We use timing of farming activities as proxy variables for these "gender influenced" variables. Specifically, we use whether planting and weeding operations are performed in a timely manner as additional explanatory variables for yield. The timing of planting is a dummy variable that indicates whether the rice field was planted late relative to the rest of the sample. The weeding variable is measured as the number of days between planting and weeding, and is expected to be negatively correlated with yield.

Table 4: OLS Results, Extended Model

Variable	Estimate	t-ratio
Labor (man hours/ha)	3.478**	3.532
Number of irrigations	117.86**	4.218
Number of days field was guarded against birds	39.826*	2.644
Split fertilizer application (1=more than 1 application)	400.63	1.529
Total farming knowledge score	0.0114	0.811
Dummy Variable for Perimeter (1=1st perimeter)	-652.66	-1.725
Dummy Variable for the Gender of the Farmer (1=male)	753.68**	2.835
Interaction term: late planting and gender	-0.0076*	-0.079
Interaction term: weeding timing and gender	-1333.7**	-3.228
Constant	361.21	1.092

Dependent variable= rice yield, kg per hectare. N=135 R^2=.41 R^2 adjusted=.38
**significant at 1% level, * significant at 5% level

The results of our *extended model* are presented in Table 4 and show that the total amount of labor, number of irrigations and number of days the field was guarded against birds continue to have a high and positive significant impact on yield. The gender dummy variable becomes more significant (and nearly doubles in value) indicating that female producers achieve yields that are on average 852 kg/ha (27%) lower than those for men. This is independent of the large and significant negative impact of late weeding, which may also be intermediate outcome associated with gender.

Comparison of Basic and Extended Model

To test whether the addition of the two timing variables ("gender influenced" variables) help to explain the variation in the dependent variable we conduct a joint significance test for this subset of regression coefficients. Using the null hypothesis that the timing coefficients are jointly equal to zero $\delta_q=0$, the appropriate F statistic is[2]: $F_{q,N-k} = F_{2,125} = 9{,}67$ (q=number of timing variables, k=total number of coefficients in the extended model, and N=number of observations). The F statistic exceeds the critical value of the F distribution at 5% level, and we therefore reject the null hypothesis of identical timing of farm operations for male and female farm operators.

Considering the nature of ε in the above *extended model*, it is generally argued that ε represents the large set of unobserved inputs into the production process. If y, X and D are observed, and ε is uncorrelated with X and D, then ordinary least squares (OLS) estimates of the parameters in our extended model are consistent. However, it may be reasonable to believe that since some X are chosen by the farmer–in particular the timing and irrigation variables--they are not randomly allocated across plots. So the error term will be highly correlated with D, and OLS parameters estimates will be biased. An appropriate estimation approach would be to use instruments which influence factor allocation decisions and yet are uncorrelated with ε.

In our *endogenous input choice model,* we try to use instrumental variables which are both highly correlated with the independent variable and at the same time uncorrelated with the error term

[2] See p. 117 in Pindyck, R. and D. Rubinfield.

in the equation. We chose a dummy variable representing whether the respondent belong to a noble caste or not as an instrument for number of irrigations and number of times the respondent was on cooking duty as an instrument for late weeding of the plot. If the farmer is a member of a noble caste, it is reasonable to assume that he/she enjoys preferential rights to irrigation water allocation. If a woman has more cooking responsibilities (as measured by the number of cooking duties she must do each week) she is likely to be have an adverse impact on the timing of her farming activities.

The results of the *endogenous input choice model* are presented in Table 5. Hypothesis testing in instrumental variable estimation is complicated by the unknown distributions, making the usual t and F tests invalid. To test whether the bias due to simultaneity is significantly different from zero, we need to bootstrap the standard error of the bias and then use the standard errors obtained from the bootstrap procedure to calculate the relevant t statistics. The results show that the total amount of labor, number of irrigations and number of days the field was guarded against birds continue to have a high and positive significant impact on yield and the late weeding continues to have negative significant impact on yield. The gender dummy variable has a significant impact on yield, indicating that female producers achieve yields that are on average 544 kg/ha (20%) lower than those for men.

Table 5: 2SLS Results, Endogenous Input Choice Model

Variable	Estimate	t-ratio[a]
Labor (man hours/ha)	3.641**	3.695
Number of irrigations	89.667**	3.073
Number of days field was guarded against birds	39.834*	2.637
Split fertilizer application (1=more than 1 application)	230.25	0.869
Total farming knowledge score	0.0122	0.879
Dummy Variable for Perimeter (1=1st perimeter)	-296.62	-0.769
Dummy Variable for the Gender of the Farmer (1=male)	543.61*	2.176
Dummy variable for planting late (1=late planting)	-0.0077	-0.798
Dummy variable for weeding late (1=late weeding)	-1728.3**	-5.975
Constant	760.81*	2.246

Dependent variable= rice yield, kg per hectare. Instrumental variables are: noble caste, number of cooking duties a week. N=135, R^2 = .37, R^2 adjusted=.32

[a]The standard errors for coefficients are generated by bootstrapping the sample set and generating 1000 new samples by replacement.

Conclusions

The inclusion of timing variables to the estimation has significant power in explaining the yield differences between male and female farmers, and hence collecting this type of data can be justified on the basis of improved estimation. The next step in the analysis (which is beyond the scope of this paper) would be to calculate the economic cost of bad decisions/policies that would result from not including the timing variables into male-female productivity analysis and compare that to the cost of collecting the necessary data.

Proper treatment of endogeneity in the estimation of male-female productivity differences still remains problematic. Our results show that there is a bias against women in timely weeding and guarding the fields against birds. We were successful in finding an instrument (cooking duty) for late weeding but failed to find a proper instrument for guarding against bird damage. Since the guarding against birds is done by children, the number of children one has was a logical choice for an instrument but this was not highly correlated with the independent variable and at the same time uncorrelated with the error term in the equation. We also identified irrigation as another endogenously determined variable in our model but this is not necessarily a gender-differentiated variable but a user-differentiated variable, where one's social status determines access to the resource. The difference (and importance) of user-differentiated and gender-differentiated variables should not be overlooked in the attempts to better explain male-female productivity differences.

References:

Dingkuhn, Michael. "Yield constraints in a Sahelian irrigation scheme." WARDA Annual Report 1992. West Africa Rice Development Association.

Pindyck, R. and D. Rubinfield. *Econometric Models and Economic Forecasts.* New York: McGraw-Hill Book Company. 1981.

Quisumbing, Agnes. "Male-Female Differences in Agricultural Productivity: Methodological Issues and Empirical Evidence." *World Development*, Vol. 24, No.10, pp. 1579-1595, 1996.