A SUBJECTIVE PRODUCTION FUNCTION FOR WHEAT IN THE PAMPEAN REGION (ARGENTINA)

L. FRANK[1] and M. DEMARCHI[2]

Recibido: 28/03/11
Aceptado: 03/08/11

SUMMARY

The paper presents a production function based on yield expectations from a sample of qualified informants of the Pampean region. The findings confirm previous estimates of the marginal productivity of nitrogen and phosphorus, but disagree on the marginal productivity of other factors. The paper introduces for the first time variables related to weeds, insects and fungi control in a local production function. The discussion at the end focuses on the consistency of the function with neoclassical assumptions about production functions.

Key words. Subjective expectations, production function, wheat.

UNA FUNCIÓN DE PRODUCCIÓN SUBJETIVA PARA TRIGO EN LA REGIÓN PAMPEANA (ARGENTINA)

RESUMEN

El trabajo presenta una función de producción sobre la base de expectativas de rendimiento de una muestra de referentes calificados de la región pampeana. Los resultados confirman estimaciones previas sobre las productividades marginales de nitrógeno y fosforo, pero discrepa respecto de las productividades marginales de otros factores. El trabajo también introduce por primera vez variables relacionadas con el control de malezas, insectos y hongos en una función de producción. La discusión al final se focaliza en la consistencia de esta función con los supuestos neoclásicos sobre funciones de producción.

Palabras clave. Expectativas subjetivas, función de producción, trigo.

INTRODUCTION

The production function is a mathematical expression that relates product and inputs according to the current state of knowledge about the production process. Identifying this function is essential to determine the optimal level of resource allocation. Therefore a great experimental effort has been oriented to clarify the relationship between product and the major inputs for grain production. It is notable, in particular, the effort done to understand the effects of fertilization on yield. Álvarez (2007, p. 91-119), for example, presents a comprehensive review of experimental results on fertilization of wheat and compiles more than ten technology functions that explain the yield of this crop in different areas of the Pampean region. These functions, however, are limited because they only consider one or two factors to explain the production. Few studies have aimed to study the impact of multiple factors (simultaneously) on grain yields, in general, and wheat yield (see e.g. Bono and Álvarez 2008, and Álvarez 2009), in particular.

1 Departamento de Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía UBA. Av. San Martín 4453, (C1417DSE) Buenos Aires, Argentina. Tel.: 54-11-4524-8077, e-mail: lfrank@agro.uba.ar.
2 Student, Licenciatura en Economía y Administración Agrarias.

An alternative source of information, which has been barely studied, is the experience of farmers and farm managers. This experience is crucial in the allocation of resources in real production situations and, therefore, it is essential to recover the production function implicit behind business decisions. This experience is also useful as an alternative source of information when experimental data are unavailable or to supplement the experimental data when these are scarce. It is well known, for example, that estimators which use *a priori* information are considerably more efficient than ordinary estimators. However, incorporating this type of information is possible only if it is consistent with the experimental data. Moreover, it is necessary that both sources are consistent with the economic theory (Chambers 1994) to obtain estimates not only statistically well-behaved, but also economically «rational». The aim of this paper is to retrieve the experience of a sample of farmers and farm managers (that we shall call qualified informants), and to check its consistency with experimental data and the theory of production functions.

OBJECTIVES

The specific objective of the paper is to propose a general predictive function of wheat yield after the experience of qualified informants, and compare this function with a previous one proposed by Frank (2011) after experimental data. Through the comparison we intend to verify:

a) The «real» production function that arises from subjective expectations is consistent with experimental findings. That is, both sources of information are complementary and lead to similar output elasticities for the major factors influencing yield.
b) The expectations of qualified informants are rational in the sense that they satisfy the neoclassical assumptions about production functions and, therefore, can be represented by functions developed by the economic theory.

Therefore, we depart from the following assumptions:

A.1 There is a single production function which is «true» (but unknown) and underlies both the experimental information and subjective expectations. The yield expectations function is «unique», that is, it is the same in all areas of the Pampa and its parameters are crop-specific and independent from the environment.

A.2 The «true» production function (although analytically complex) is linear in the parameters, and includes as relevant variables N, P and chemicals, as well as rainfall, temperature and soil texture.

A.3 Yield expectations of producers and consultants are unbiased. They are formed mainly from personal experience, but also through public or private technical information.

The paper contains numerous technical details that could easily divert the reader from the main idea which is to present a technological function (estimated from subjective expectations of yield) and compare its coefficients with those arising from experimental evidence as well as those expected according to the economic theory. For this reason we suggest the reader to skip, in a first reading, the subsection on parameter estimation, going directly to the results and discussion, and to return in further readings to the estimation details.

In the forthcoming section we present a production function whose main characteristic is to consider both productive and environmental factors, the former subject to the restrictions imposed by economic theory. In this context, we propose a logarithmic relationship between yield and nitrogen available in soil but penalized by phosphorus deficit, and a log-linear relationship with a saturation break to relate yield and control of weeds, pests and fungi. The type of tillage operates in the model as a simple scale variable without interacting with the others. Moreover, we include precipitation, temperature and soil textural class as environmental variables. At the end of the paper we compare the coefficients associated with all these variables against estimates based on agricultural experiments and discuss their meaning in the context of the theory of production functions.

MATERIALS AND METHODS

The study had four stages: (a) design and distribution of a questionnaire on yield expectations, (b) estimation of nitrate at sowing (which we justify below), (c) selection of a technological function, and (d) parameter estimation and hypothesis testing. We describe each stage in detail below.

The survey of yield expectations

We designed a questionnaire on the expected yields under various productive and environmental «scenarios». Each scenario is a possible combination of production factors, whether controllable or not. The specification of scenarios was achieved after preliminary enquiries to agronomists directly involved in production. Scenarios that, although possible, are hardly experienced were excluded from the questionnaire, *e.g.* extreme drought or application of chemicals in toxic doses. We designed two forms (that we shall call I and II) with 126 scenarios each. Additionally, we included a third blank form for the respondent to answer on production or environmental conditions different than those provided in forms I and II. However, this form was rarely used.

Forms were distributed through students of the course of Econometrics (editions 2009 and 2010) and by the author to graduates of the Faculty of Agronomy, UBA[3]. Among the questionnaires distributed in 2009 (mainly by students) form I was predominantly used, while among the questionnaires distributed in 2010 (mainly by the author) form II was the prevailing. For this reason it was observed that the geographical coverage of the survey in 2009 was wider than that of 2010. In this last year most responses were from northern Buenos Aires and the province of Entre Ríos. We received 51 forms with a total of 6,609 responses. The entire database of responses (excluding the identification of the informant) is available to interested parties upon request.

Estimation of nitrates at sowing

The nutritional variables that really explain wheat yield are the levels of nitrogen and phosphorus available for the crop (Álvarez 2007), *i.e.* the sum of each nutrient present in the soil plus that provided by the fertilizer. To estimate available nitrogen and phosphorus levels we compiled technical reports from the on-line repositories of INTA (http://www.inta.gov.ar) and IPNI (http://www.ipni.net). These reports covered the period 1996-2010. In total we collected 129 records of nitrogen as nitrate ($N\text{-}NO_3^-$) and 94 records of extractable phosphorus. Because some accounts reported $N\text{-}NO_3$ in the first 20 cm of the soil profile while others reported $N\text{-}NO_3^-$ up to 60 cm, we chose to standardize all the observations at 60 cm using a conversion function. To estimate this function we fit a simple linear regression, where records at 60 cm (in $kg.ha^{-1}$) were the dependent variable and records at 20 cm (in ppm) were the independent variable. To do so we used data from 30 reports with records at both depths. Thus 40 values out of the 129 listed above were estimated in this fashion. It should be noted that some «records» were actually averages of observations from the same area, and that is why the conversion function was computed following a two steps generalized least squares (GLS) procedure. Extractable P records did not require conversion as all the records were at 20 cm depth in units of [ppm]. Recall that [ppm] units may be transformed into [$kg.ha^{-1}$] of phosphorous (P-P_2O_5) by multiplying by a factor of 2.67 (see Álvarez 2007) approximately.

3 It can be inferred that the population of this sample are the managers of medium-sized to large farms in the area of influence of the Faculty of Agronomy of the UBA.

Production function

We departed from the function proposed by Frank (2011) based on experimental data compiled by Álvarez (2007, p. 91-119). This function is an adaptation of the «generalized power production function» (De Janvry 1972b), in turn inspired by Argentinean (De Janvry 1972a) data. The function is consistent with the neoclassical assumptions about production functions. Briefly, the function has two components, a productive component $f(\mathbf{x})$, and an environmental component $g(\mathbf{x})$. The first component considers as main input available nitrogen (N_d) and as secondary input available P, while the second component considers rainfall, temperature and soil characteristics. The type of tillage enters the function as a scale variable associated with the first component. We added two extra variables to the function associated with the control of weeds, insects and fungi as shown below. We write the expected-yield technological function in analytical form as

$$\ln(y_i) = (c_0+c_1) + f(x_{i2}, \ldots, x_{i6}) + g(x_{i7},\ldots, x_{i11}) + \varepsilon_i \text{ with } \varepsilon_i\sim N(0,\sigma_i^2),$$

where

$$f(\mathbf{x}_i) = \alpha_2 \ln(x_{i2}) + \alpha_3 (z-z_0) \ln(x_{i2})\, \delta_{z\leq z(0)} + \alpha_4 x_{i4} + \alpha_5 (x_{i4}-x_4^*)\, \delta_{x(4)>x(4)^*} + \alpha_6\, \delta_{h=\mathrm{SD}}$$

and

$$g(\mathbf{x}_i) = \alpha_7 x_{i7} + \alpha_8 (x_{i7}-x_7^*)\, \delta_{x(7)>x(7)^*} + \alpha_9 x_{i9} + \alpha_{10} x_{i10} + \alpha_{11} x_{i11} \quad (1)$$

The meaning of each variable in matrix format is as follows:

- $\mathbf{y}$ is the vector of wheat yields in logarithmic scale. Each element of $\mathbf{y}$ is $\mathbf{y}_i = \sum_{r=1,m} \ln(y_{ir})\, m_i^{-1}$ where y_{ir} in [kg.ha^{-1}] is the expected yield under the i-th scenario.
- α_j for all $j = \{1,..., k\}$ are fixed (but unknown) parameters of the wheat crop.
- $\mathbf{x}_1$ = **1** is the variable associated with the constant $(c_0 + c_1)$, where c_0 is the intercept associated with $f(\mathbf{x}_i)$ and c_1 is the intercept associated with $g(\mathbf{x}_i)$.
- $\mathbf{x}_2$ is the level of N_d down to 60 cm depth (in kg.ha^{-1}) and in logarithmic scale; N_d is the sum of $N\text{-}NO_3^-$ present in the soil (see table 2) and $N\text{-}NO_3^-$ from the fertilization.
- $\mathbf{x}_3$ $= (\mathbf{z}-z_0\mathbf{1})\circ\ln(\mathbf{x}_2)\circ\boldsymbol{\delta}_{z\leq z(0)}$ is the difference between the level of $P\text{-}P_2O_5$ in [kg.ha^{-1}] in the first 20 cm of depth and the critical level z_0 = 15 ppm (≈ 40 kg) multiplied by $\ln(x_2)\ \delta_{z\leq z(0)}$; $\delta_{z\leq z(0)}$ is a Kronecker delta that equals 1 if $z \leq z_0$ or 0 otherwise. The symbol «o» indicates the Hadamard product[4].
- $\mathbf{x}_4$ is the amount of «sprays» of chemicals against weeds, insects and fungi in [units.cycle^{-1}]. Although it is a discrete variable, encoding x_4 from scenarios where sprays were defined by range resulted in real values.
- $\mathbf{x}_5$ $= (\mathbf{x}_4-x_4^*\mathbf{1})\circ\boldsymbol{\delta}_{x(4)\geq x(4)^*}$, where $x_4^* = 2$ [units.cycle^{-1}], and x_5 is a variable equal to $(x_4 - x_4^*)$ if more than two chemicals are sprayed or zero otherwise. We set up x_4^* through enquiries to agronomists and by graphical inspection of the response.
- $\mathbf{x}_6$ $= \boldsymbol{\delta}_{h=\mathrm{SD}}$ is a variable indicating type of tillage: $\delta_{h=\mathrm{SD}}$ equals 1 if it is SD or 0 otherwise. Recall that Frank's (2011) original function considered $x = \delta_{h=\mathrm{SD}}2^{-1}+(1-\delta_{e=\mathrm{LC}})2^{-1}$ to admit an intermediate value for records not mentioning the type of tillage.
- $\mathbf{x}_7$ is the total rainfall during the crop cycle [mm. cycle^{-1}].
- $\mathbf{x}_8$ $= (\mathbf{x}_7-x_7^*\mathbf{1})\circ\boldsymbol{\delta}_{x7\geq x7^*}$ where $x_7^* = 400$ [mm. cycle^{-1}], and x_8 is a variable equal to $(x_7-x_7^*)$ if the rainfall exceeds the critical level x_7^* or zero otherwise. We established this value as a temporal proportion of the critical level given by Frank (2011) and assuming symmetry in the annual distribution of rainfall.

4 A Kronecker delta $\delta_{i=j}$ is a binary variable that equals 1 when $i=j$ or 0 otherwise. The Hadamard product is defined as the element by element product of two matrices.

$\mathbf{x}_9$ is the average temperature during the crop cycle in [°C]; this variable was included later to fully reproduce Frank's original function. The inclusion of $\mathbf{x}_9$ determined that the total amount of scenarios largely exceeded the 252. The average temperatures proceeded from SNIH (2001) cartography and laid within the range 14.5-18.0 °C.

$\mathbf{x}_{10} = \boldsymbol{\delta}_l$ where δ_l is a binary variable equal to 1 if the soil texture is loam or silt-loam or zero otherwise.

$\mathbf{x}_{11} = \boldsymbol{\delta}_r$ where δ_r indicates clay-loam or clay textures. We omit δ for sandy or sandy-loam textures to avoid any linear dependence with columns $\mathbf{x}_{10}$ and $\mathbf{x}_{11}$.

$\boldsymbol{\varepsilon}$ is the error term and includes all the variables that were omitted from the model for some reason. Each e_i is the average of $\{\varepsilon_{i1},\ldots,\varepsilon_{ih},\ldots,\varepsilon_{im}\}$ errors distributed ε_{ih}~$N(0,\sigma^2)$. The variance of ε_i is $var(\varepsilon_i) = \sigma^2 m_i^{-1}$. We assume that $cov(\varepsilon_i,\varepsilon_{i'}) = 0$.

The econometric model associated with this function is

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon} \text{ where } \boldsymbol{\varepsilon}\sim N(\mathbf{0},\sigma^2\boldsymbol{\Omega})$$

where $\mathbf{y}$ is a vector of dimension $n\times 1$, $\mathbf{X} = [\mathbf{X}_1|\mathbf{X}_2]$ is a matrix of known constants of dimension $n\times k$ (which for didactic reasons we decompose into a matrix $\mathbf{X}_1$ of production variables and another matrix $\mathbf{X}_2$ of environmental variables, respectively), $\boldsymbol{\beta} = [\boldsymbol{\beta}_1'|\boldsymbol{\beta}_2']'$ is a vector of fixed but unknown parameters of dimension $k\times 1$, and $\boldsymbol{\varepsilon}$ is a vector of $n\times 1$ unobservable random variables. $\sigma^2\boldsymbol{\Omega}$ is a diagonal and positive definite matrix. Each diagonal element of $\boldsymbol{\Omega}$, ω_{ij} is defined

$$\ln(\omega_{ij}) = \ln(\theta)\ \delta_{s=\mathrm{I}} - \ln(m_i) \text{ for all } i = j, \text{ or } \omega_{ij} = 0 \text{ for all } i \neq j, \tag{2}$$

where θ is a scale factor and $\delta_{s=\mathrm{I}}$ is a Kronecker delta that equals 1 if the observation comes from form I and 0 otherwise.

Parameter estimation

To estimate the parameters we essentially follow the same steps of Frank (2011) with a slight adaptation regarding the estimate for $\boldsymbol{\Omega}$. They are:

i. We computed the condition number $\kappa(\mathbf{X})$ to detect possible linear dependence relationships between the regressors[5]. However, since the beginning of the study we decided to retain all the variables of the original function, unless $\kappa(\mathbf{X}) > 100$, *i.e.* $\mathbf{X}$ exhibited severe multicollinearity (Judge *et al.*, p. 902, Greene, 2006, p. 59-61).

ii. We estimated the parameters of the model by ordinary least squares (OLS). Recall that the expression of this estimator is $\mathbf{b}_{\mathrm{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

iii. We computed the residual $\mathbf{e} = \mathbf{y}\text{-}\mathbf{X}\mathbf{b}_{\mathrm{OLS}}$ and we estimated matrix $\boldsymbol{\Omega}$ through the auxiliary regression

$$\ln(\varepsilon_i^2) = \ln(\sigma^2) + \ln(m_i^{-1}) + \ln(\theta)\ \delta_{s=\mathrm{I}} + v_i \text{ where } v_i\sim N(0,\sigma_i^2). \tag{3}$$

where ε_i^2 is estimated e_i^2. Then we replaced θ in (2) with the estimate calculated in (3). Recall that $\boldsymbol{\Omega}$ is a diagonal matrix, so it is only necessary to estimate n elements ω_{ij} for all $i = j$.

5 Although we tried to define scenarios that combine all possible levels of production and environmental factors, the exclusion of unrealistic situations could have generated some level of multicollinearity among the columns of **X**.

iv. We estimated again the vector of parameters, but this time by «feasible» generalized least squares (GLS) or FGLS. The GLS estimator is $\mathbf{b}_{GLS} = (\mathbf{X'\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X'\Omega}^{-1}\mathbf{y}$ and the FGLS estimator is constructed replacing $\mathbf{\Omega}$ with its estimator obtained in the previous step. For computational ease we actually transformed matrix **X** and vector **y** dividing each row of them by the estimated $\omega_{ij}^{1/2}$ and proceeded as in OLS estimation. This procedure is exactly the same as the matrix operations given before.

v. We estimated the variances of $\mathbf{b}_{FLS}$ and $\mathbf{b}_{FGLS}$. Their expressions are $var(\mathbf{b}_{LS}) = \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'\Omega X}(\mathbf{X'X})^{-1}$ and $var(\mathbf{b}_{GLS}) = \sigma^2(\mathbf{X'\Omega}^{-1}\mathbf{X})^{-1}$, respectively. The LS (least squares) estimator is White's heteroskedastic estimator and its use is recommended (Frank 2010) for small samples and ill-conditioned matrices **X**.

vi. Finally, we computed the adjusted R^2 statistic and the t statistics to test $\mathbf{ß} = \mathbf{0}$. We checked normality of the residuals by the Jarque-Bera normality test on $\mathbf{e}_{LS} = \mathbf{y}\text{-}\mathbf{Xb}_{FGLS}$ and $\mathbf{e}_{GLS} = \mathbf{y}\text{-}\mathbf{Xb}_{FLS}$.

vii. In addition, we tested two hypotheses on the slopes of the broken lines relating yield to pest control and rainfall. Specifically, we tested for a linear plateau relationship through the hypothesis $ß_j + ß_{j+1} = 0$ for $j = \{4,7\}$. Therefore, we considered the linear system $\mathbf{r = Rß}$, where **R** is a full row rank matrix of dimension $q\times k$ and **r** is a vector of dimension $q\times 1$. In our case we only had $q = 2$ hypothesis. Then we calculated the statistic

$$\lambda = (\mathbf{Rb\text{-}r})'(\mathbf{R}\varepsilon^2\mathbf{R'})^{-1}(\mathbf{Rb\text{-}r})/q$$

where ε^2 is $var(\mathbf{b}_{GLS})$. This statistic is distributed exactly $\lambda \sim \chi^2_{(q)}$ if $\varepsilon_i \sim N(0,\sigma_i^2)$ and it is possible to prove that even if ε_i is not normally distributed, but n is «sufficiently» large, then $\lambda^*(s^2\mathbf{\Omega}^*) \rightarrow^d \chi^2_{(q)}$.

viii. For reasons that will become evident in the discussion we also computed **b** through two other estimators, Theil's (1963) estimator and the LAD (Least Absolute Deviation) estimator. Theil's estimator combines current information with *a priori* information expressible as $\mathbf{r = Rß + \nu}$, where $\boldsymbol{\nu} \sim N(0, \sigma_i^2)$. The expression of the Theil estimator is:

$$\mathbf{b}_{Theil} = (\mathbf{X'\Omega}^{-1}\mathbf{X}+w\mathbf{R'\psi}^{-1}\mathbf{R})^{-1}(\mathbf{X'\Omega}^{-1}\mathbf{y}+w\mathbf{R'\psi}^{-1}\mathbf{r}),$$

and its variance is

$$var(\mathbf{b}_{Theil}) = \sigma_\varepsilon^2(\mathbf{X'\Omega}^{-1}\mathbf{X}+w\mathbf{R'\psi}^{-1}\mathbf{R})^{-1},$$

where $w = \sigma_\varepsilon^2\sigma_\nu^{-2}$. The LAD estimator is a robust estimator that minimizes the sum of absolute deviations rather than the error sum of squares (Pynnönen & Salmi 1994, Dasgupta 2004) and has the advantage of weighing less extreme observations. The LAD estimator is the solution of a linear programming problem.

All calculations were performed in the free software matrix language Euler Math Toolbox v10.1 developed by René Grothmann (2010), Associate Professor of Katholische Universität Eichstätt (Germany). The codes are available upon request.

RESULTS

Table 1 presents the regression coefficients and other statistics of the conversion function of $N\text{-}NO_3^-$ at 20 cm to 60 cm. The adjusted R^2 coefficients were 0.6284 and 0.6276 for the OLS and GLS estimation, respectively. Table 2 presents the medians and extreme values of the sample classified by wheat sub-regions. You can see that the level of $N\text{-}NO_3^-$ is extremely variable, even within each sub-region.

TABLE 1. **Regression coefficients of the function that converts N-NO$_3^-$ at 20 cm [ppm] into N-NO$_3^-$ at 60 cm [kg.ha^{-1}]**

Estimator	b_j	$s_{b(j)}$	t stat.	$P(\|T\|>t)$
OLS	10.916	7.347	1.486	0.140
	3.878	0.548	7.074	—
GLS	9.379	4.910	1.910	0.058
	4.006	0.414	9.670	—

TABLE 2. **Median of N-NO$_3^-$ [kg.ha^{-1}] and extreme values classified by wheat sub-region**

Subreg.		N-NO$_3^-$				P		
	min.	median	max.	n_N	min.	median	max.	n_P
I	25.9	56.4	135.6	18	5.6	26.2	65.0	8
IIN	30.0	72.0	169.6	27	7.8	14.0	38.0	22
IIS	27.0	55.9	97.5	25	4.0	17.0	71.0	25
III	46.6	48.6	50.6	2	13.0	17.3	21.6	2
IV	38.2	48.0	70.6	7	13.5	19.7	25.9	9
VN	40.6	46.4	64.7	4	5.0	22.5	37.0	4
VS	21.0	62.6	237.7	46	6.1	17.5	51.6	24
Gral.	21.0	57.2	237.7	129	4.0	17.1	71.0	94

Following the estimation protocol, the condition number of **X** was 99.66, revealing moderate to strong linear dependencies among the columns of **X**. The estimated value of ln(θ) was 0.8935. Tables 3 and 4 present the regression coefficients obtained by LS and GLS, as well as their deviations and t-statistics[6]. The adjusted R^2 coefficients were $R^2 = 0.5955$ and $R^2 = 0.8852$, respectively. The λ statistic allowed us to reject the null hypothesis $\alpha_j + \alpha_{j+1} = 0$ for $j = \{4,7\}$. However, further sequential tests only rejected the hypothesis $\alpha_7 + \alpha_8 = 0$. According to the Jarque-Bera normality test we rejected normality in the residuals with type I error probability of 0.05 and 0.01.

TABLE 3. **Regression coefficients obtained by FLS**

Associated variable	b_j	$s(b_j)$	t stat.	$P(\|T\|>t)$
$\mathbf{x}_1$ = intercept	3.3849	0.2096	16.1515	—
$\mathbf{x}_2$ = ln(N$_d$)	0.3210	0.0279	11.4991	—
$\mathbf{x}_3$ = P deficit × ln(x_2) $\delta_{z\leq z(0)}$	-0.0256	0.0045	-5.7241	—
$\mathbf{x}_4$ = chemical controls	0.1061	0.0120	8.8255	—
$\mathbf{x}_5$ = chemical controls > 2	-0.0782	0.0286	-2.7329	0.0065
$\mathbf{x}_6$ = zero tillage	0.0886	0.0155	5.7294	—
$\mathbf{x}_7$ = rainfall	0.0022	0.0001	14.8642	—
$\mathbf{x}_8$ = rainfall >400 mm	-0.0015	0.0002	-7.4600	—
$\mathbf{x}_9$ = temperature	-0.1542	0.0098	-15.6592	—
$\mathbf{x}_{10}$ = loam or silt-loam soil	-0.0471	0.0197	-2.3908	0.0171
$\mathbf{x}_{11}$ = clay-loam or clay soil	0.0288	0.0180	1.5973	0.1107

6 We do not report values of $P(|T| > t) < 10^{-4}$.

TABLE 4**. Regression coefficients obtained by FGLS**

Associated variable	b_j	$s(b_j)$	t stat.	$P(\|T\|>t)$
$\mathbf{x}_1$ = intercept	3.3373	0.2585	12.9125	—
$\mathbf{x}_2 = \ln(\mathrm{N_d})$	0.3522	0.0335	10.5222	—
$\mathbf{x}_3 = \mathrm{P\ deficit} \times \ln(x_2)\ \delta_{z \le z(0)}$	-0.0168	0.0053	-3.1580	0.0017
$\mathbf{x}_4$ = chemical controls	0.0851	0.0143	5.9309	—
$\mathbf{x}_5$ = chemical controls > 2	-0.0464	0.0342	-1.3556	0.1757
$\mathbf{x}_6$ = zero tillage	0.0870	0.0185	4.7092	—
$\mathbf{x}_7$ = rainfall	0.0021	0.0002	11.5405	—
$\mathbf{x}_8$ = rainfall >400 mm	-0.0014	0.0003	-5.5724	—
$\mathbf{x}_9$ = temperature	-0.1576	0.0124	-12.7483	—
$\mathbf{x}_{10}$ = loam or silt-loam soil	-0.0674	0.0236	-2.8616	0.0044
$\mathbf{x}_{11}$ = clay-loam or clay soil	0.0462	0.0216	2.1408	0.0327

Table 5 presents estimates of the same parameters but incorporating an *a priori* estimation of α_3 and also a binary variable, indicative of a response detected as atypical. The estimator labeled Theil I assumes that $\sigma_\varepsilon^2 \approx \sigma_v^2$ while the estimator labeled Theil II considers the estimates of s_e^2 and s_b^2 known from the FGLS regression and from Frank (2011), respectively. The last column of table 5 shows the elasticities of $\mathrm{N_d}$, P, chemicals and rainfall and the semi-elasticities of temperature and soil texture in a situation of 100 kg $\mathrm{N_d}$, deficit of 5 ppm (≈13, 3 kg) P, 3 sprays of chemicals and 450 mm rainfall.

TABLE 5**. Regression coefficients obtained by OLS, Theil and LAD and elasticities after the LAD estimator.**

Associated variable	Theil I	Theil II	LAD	Elasticity
$\mathbf{x}_1$ = intercept	3.3457	3.3447	3.6528	—
$\mathbf{x}_2 = \ln(\mathrm{N_d})$	0.3637	0.3642	0.3326	0.31127
$\mathbf{x}_3 = \mathrm{P\ deficit} \times \ln(x_2)\ \delta_{z \le z(0)}$	0.0016	0.0016	0.0016	0.19649
$\mathbf{x}_4$ = chemical controls	0.0552	0.0552	0.0572	0.13020
$\mathbf{x}_5$ = chemical controls > 2	-0.0288	-0.0283	-0.0138	—
$\mathbf{x}_6$ = zero tillage	0.0752	0.0753	0.0714	0.07140
$\mathbf{x}_7$ = rainfall	0.0018	0.0018	0.0017	0.31500
$\mathbf{x}_8$ = rainfall >400 mm	-0.0012	-0.0012	-0.0010	—
$\mathbf{x}_9$ = temperature	-0.1550	-0.1551	-0.1643	-0.16430
$\mathbf{x}_{10}$ = sand and sand-loam soil	0.0943	0.0944	0.0901	—
$\mathbf{x}_{11}$ = clay-loam or clay soil	0.0669	0.0683	0.1435	0.14350
$\mathbf{x}_{12}$ = outlier	-0.7468	-0.7456	-0.7925	—

DISCUSSION AND CONCLUSION

In view of the R^2 coefficients, the t tests and the matching order of the estimates compared to their experimental counterparts, we conclude that function (1) is adequate to explain the yields expected by qualified informants and also that it is consistent with previous findings from agricultural experiments. This means that qualified informants are able to predict wheat yield accurately and that the underlying function behind their expectations is similar to that arising from experiments. However, besides consistency between both sources of information we also require the function to be compatible with the theoretical assumptions about production functions, a point which we want to discuss below.

In (1) α_2 is the elasticity of output with respect to N_d for non-limiting $P\text{-}P_2O_5$ levels. For limiting levels of $P\text{-}P_2O_5$ ($z_0 < 15$ ppm) the elasticity appears penalized proportionally to the deficit of $P\text{-}P_2O_5$ through the term $\alpha_3(z\text{-}z_0)$. Logically, the penalty factor is positive since the elasticity $\alpha_2 + \alpha_3(z\text{-}z_0)\,\delta_{z \leq z(0)} > 0$ and $\alpha_2 > 0$ for $0 \leq z \leq +\infty$. This implies, in turn, that $|z\text{-}z_0| < \alpha_2\alpha_3^{-1}$ and, therefore, that $\alpha_3 > 0$. However, the results show that $\alpha_3 < 0$, contrary to the theory. This result is absurd because it would imply that the greater the deficit of P, the greater the elasticity of N_d. This inconsistency could be explained in two ways: (a) because «atypical» observations contaminated the data, resulting in a bias in the estimate, and (b) because $P\text{-}P_2O_5$ deficit was estimated incorrectly, either because the levels of z in Table 2 are wrong, or because the critical level for z_0 suggested in the experimental literature differs from that perceived by the informants.

To verify (a) we computed the coefficients using the LAD estimator in order to reduce the weighting of extreme observations. This new estimate shows that $b_{3,\mathrm{LAD}} \approx -0{,}008 >> b_{3,\mathrm{LS/GLS}} \approx -0{,}020$ suggesting that indeed some extreme points altered the initial estimate. Unfortunately, no statistics have been developed to allow testing hypotheses with the LAD estimator, but the rule of thumb $|b_{3,\mathrm{LAD}}\, s(b_3)^{-1}| < 2$ suggests that $\alpha_3 \approx 0$ can not be rejected. In view of this result, we performed a thorough inspection of the database through which we identified six extreme values in one of the questionnaires. We classified these answers with a dummy variable which we called x_{12}[7]. To verify (b) we tested the hypothesis H_0: $d = z\text{-}z_0 = 0$ for each sub-region of Table 2 and rejected the null in all cases. Then we conclude that the analyzed data do not provide enough evidence of $P\text{-}P_2O_5$ deficiency in any of the sub-regions and therefore $\mathbf{x}_3 = \mathbf{0}$ should be excluded from the model or at least supplemented with the prior information $ß_3 = -1.55 \times 10^{-3} + \nu$ with $\nu \sim N(0,\sigma_i^2)$ provided by previous estimates[8]. Table 5 presents the estimates obtained by the latter alternative using Theil's (1963) estimator, and adding the variable x_{12} for outliers. You can see that the elasticity of N_d from the expectations functions is slightly (but significantly) higher than that calculated previously by Frank (0.32-0.36 *vs.* 0.24-0.25, respectively) based on experimental data, but within the range 0.16-0.42 arising from the literature (Álvarez 2007).

The signs of α_j and α_{j+1} for $j = \{4,7\}$ (see Tables 3, 4 and 5) are coherent with the theory, as we know that $\alpha_j > 0$ in the economically feasible region and $\alpha_{j+1} < 0$ to ensure decreasing marginal productivity. Furthermore, we can prove (see appendix) that $|\alpha_j| > |\alpha_{j+1}|$, which is also confirmed by our estimates. In the particular case of $j = 4$, the findings show that $\alpha_4 + \alpha_5 = 0$ according to the LS and GLS estimates but $\alpha_5 = 0$ according to the Theil II estimate, which in turn means that yield either reaches a plateau after the second chemical spray or remains growing at the same rate throughout the interval $1 \leq x_4 \leq 5$, provided the other variables remain constant. So far then we can only narrow α_5 in the interval $-\alpha_4 < \alpha_5$ d» 0. Furthermore we note that in the way that we defined x_4, it is a non-essential input in the neoclassical sense, since $\lim_{x(4)+\to 0} f(\mathbf{x}) \neq 0$[9]. While this situation is indeed realistic, since in practice it is verified that $y > 0$ when $x_4 = 0$, it should be noted that the variable x_4 «quantity of chemical sprays» is actually a proxy of the real variable «weeds, plagues and fungi control». However, this underlying variable also includes the control of weeds by mechanical means under LC technology, while under SD the function is not defined for $x_4 = 0$ since the technological package includes one chemical spray against weeds before sowing. Clearly, the function definition is ambiguous for $x_4 < 1$ and so for the moment we restrict x_4 to the range $1 \leq x_4 \leq 5$ until this issue is solved theoretically. According to Tables 3, 4 and 5 the semi-elasticity of x_4 lies between 0.05-0.10 depending on the estimator used.

---

7 This questionnaire had strongly decreasing returns with fertilization doses of N exceeding 70 $\mathrm{kg.ha^{-1}}$.

8 Note that the observations within sub-regions on table 2 are conspicuously erratic.

9 Here $f(\mathbf{x})$ refers to (1) without distinguishing between productive and environmental components.

Regarding rainfall, the λ_2 statistic showed that $\alpha_7 + \alpha_8 > 0$, *i.e.* that yield is monotonically increasing throughout the interval under study, contrary to experimental evidence. In addition, $\lim_{x(7)+\to 0} g(\mathbf{x}) \neq 0$. This is, without doubt, a serious drawback in the case of precipitation, as it is not realistic to assume that $y >> 0$ when $x_7 = 0$. This means that another linear function should be placed between 0 and $x_7 = 200$ mm with a break point in an unknown $x_7' < 200$. We certainly don't have yield expectations in this interval as the situation hardly ever occurs in the Pampean region. For this reason we restrict our function to the interval $200 \leq x_7 \leq 700$ mm assuming that some additional terms to explain yield in the range $0 \leq x_7 \leq 200$ mm are missing. The semi-elasticity of rainfall (17×10^{-4}-22×10^{-4}, depending on the estimator) is greater than the experimental one, but in line with that calculated by Bono and Álvarez (2008).

The regression coefficient associated with SD is considerably lower in the expectations function (0.07-0.09) than in the function arising from experimental data (0.21-0.25) and the difference between both coefficients is significant for a type I error probability of 5% and 1%. To justify this result recall that Frank (2011) defined α_6 as the product of multiple factors that contribute to yield in small amounts, that is, $\alpha_6 = \varphi_1\, \varphi_2 \ldots \varphi_p$ where $\alpha_6 > 1$ although $0 < \varphi_j << \infty$. Frank (2011) identified two of these factors: one associated with a bigger water retention capacity of soils under SD and another with the replacement of traditional genotypes with new high-performing, but more sensitive to water stress, genotypes. We propose that the discrepancy between the estimates of α_6 could be due to a third factor (let's call it φ_3) associated with inefficiencies inherent to real production situations (*e.g.* defective weeds eradication prior to sowing, mechanical failures of the seeder, pest control opportunities lost due to weather conditions, etc.). It is not possible at the moment to test this hypothesis.

The sign of temperature matches the experimental evidence, but is larger than the latter in absolute value. On the contrary, the sign of b_{11} (clay-loam to clay soil texture) coincides with that of the variable vertisols of the experimental function, but is considerably smaller. For comparative purposes we re-parameterized the variables x_{10} and x_{11}, so as to set to 0 the parameter associated with the textural class loam to silt-loam in line with the variable mollisols of the original function. Then we re-estimated the parameters using Theil's estimator and the LAD estimator. The results (Table 5) show a positive effect on the yield of sandy and clayed textural classes, unlike the experimental results, which show a negative effect of entisols-aridisols and a positive effect of vertisols. The reasons for this discrepancy are unclear. However, recall that the variable entisols-aridisols was the one that showed the highest linear dependence with other columns of **X** in Frank's (2011) study, so any result regarding this variable should be considered with caution.

AKNOWLEDGEMENTS

This paper was supported by IICA Canada and the Emerging Leaders in the Americas Program (ELAP) through one of the author's internship at the University of Manitoba.

REFERENCES

ÁLVAREZ, R. 2007. Capítulo 7: Fertilización de trigo. *In*: ÁLVAREZ, R. (*ed*.), Fertilización de cultivos de grano y pasturas. Diagnóstico y recomendación en la región Pampeana. Ed. Facultad de Agronomía UBA, pp. 91-119.

ÁLVAREZ, R. 2009. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. *European Journal of Agronomy* 30: 70-77.

BONO, A. and R. ÁLVAREZ 2008. Rendimiento de trigo en la Región Semiárida y Subhúmeda Pampeana. Un modelo predictivo de la respuesta a la fertilización nitrogenada. *Informaciones Agronómicas del Cono Sur* 41: 18-21.

CHAMBERS, R. 1994. Applied production analysis. A dual approach. Cambridge University Press.

DASGUPTA, M. and S.K. MISHRA. 2004. Least absolute deviation estimation of linear econometric models: A literature review. Available at http://mpra.ub.uni-muenchen.de/1781/ .

DE JANVRY, A. 1972a. Optimal Levels of Fertilization under Risk: the Potential for Corn and Wheat Fertilization Under alternative price Policies in Argentina. *American Journal of Agricultural Economics* 54(1): 1-10.

DE JANVRY, A. 1972b. The Generalized Power Production Function. *American Journal of Agricultural Economics* 54 (2): 234-237.

FRANK, L. 2010. Constrained Estimation with Distorted Data by the Least-Squares Criterion. *2010 Proceedings of the American Statistical Association*, Business and Economic Statistics Section, Alexandria, VA: American Statistical Association: 1926-1932. Available at https://www.amstat.org.

FRANK, L. 2011. The Wheat Production Function in the Pampa Region. Unpublished.

GREENE, W. 2006. Econometric Analysis. Prentice Hall.

GROTHMANN, R. 2010. Euler Math Toolbox v. 10.1. Available at: http://eumat.sourceforge.net.

JUDGE, G.; W. GRIFFITHS; R. CARTER HILL; H. LÜTKEPOHL and T. CHAO LEE. 1985. The Theory and Practice of Econometrics. Wiley Series in Probability and Statistics.

PYNNÖNEN, S. and T. SALMI. 1994. A Report on Least Absolute Deviation Regression with Ordinary Linear Programming. *Finnish Journal of Business Economics* 43(1): 33-49.

SNIH (SISTEMA NACIONAL DE INFORMACIÓN HÍDRICA) 2001. Mapas de precipitaciones y temperatura media anual (1965-1982). Available at: http://www.hidricosargentina.gov.ar.

THEIL, H. 1963. On the Use of Incomplete Prior Information in Regression Analysis. *Journal of the American Statistical Association* 58(302): 401-414.

APPENDIX

Following are the proofs that $\alpha_j > 0$, $\alpha_{j+1} < 0$ and $|\alpha_j| \geq |\alpha_{j+1}|$ for $j = \{4,7\}$ in expression (1):

a) $\alpha_j > 0$. We know that $x_j \geq 0$ because it is a physical quantity. If α_j were less than 0, at any point $x_j' > x_j$, $f(x_j') < f(x_j)$ which is out of the feasible economically rational area. Thus, $\alpha_j > 0$.

b) $\alpha_{j+1} < 0$. Let's consider a point x_j' such that $x_j' < x_j$ and $x_j' < x_j^*$, and a point x_j such that $x_j > x_j^*$. If both points lie in the economically feasible area, where $f(x_j) > f(x_j')$, then it is also satisfied that

$$\alpha_j x_j + \alpha_{j+1} (x_j - x_j^*) \delta_{x(j) \geq x(j)^*} > \alpha_j x_j',$$

so that

$$-\alpha_j \alpha_{j+1}^{-1} > [(x_j - x_j^*) \delta_{x(j) \geq x(j)^*}] (x_j - x_j')^{-1}.$$

But $x_j - x_j^* > 0$ and $x_j - x_j' > 0$, and by (a) we know that $\alpha_j > 0$, so $[(x_j - x_j^*) \delta_{x(j) \geq x(j)^*}] (x_j - x_j')^{-1} > 0$. This implies that

$$-\alpha_j \alpha_{j+1}^{-1} > 0 \text{ and thus } \alpha_{j+1} < 0.$$

c) $|\alpha_j| \geq |\alpha_{j+1}|$. If $f(x_j) > f(x_j')$, Then $\alpha_j (x_j - x_j') + \alpha_{j+1} (x_j - x_j') \delta_{x(j) \geq x(j)^*} > 0$, which in turn implies that

$$(\alpha_j + \alpha_{j+1} \delta_{x(j) \geq x(j)^*}) (x_j - x_j') > 0.$$

This means that

$$\alpha_j + \alpha_{j+1} \alpha_{x(j) \geq x(j)^*} > 0,$$

leading to $\alpha_j > 0$ and $|\alpha_j| > |\alpha_{j+1}|$.