Crop production function - study

Tereza Suchánková[1], Radka Bezděkovská[2]

Abstract. In general, the agricultural crops are significantly dependent on climate factors. Such variables could be not only the temperature and rainfall but also the soil moisture and the level of evaporation. Nevertheless the other factors have also an important role in the variability of crop production - classical production factors (capital, land and labor) or fertilizers and pesticides. The modeling of crop production is complex problem and needs the sophisticated approach. The aim of the paper is to study the most appropriate form of crop production function - the applied variables and mathematical form. The CES and VES production function would be the convenient tool. The estimation is based on panel data (time and Czech districts/regions) and on the outputs of hydro-meteorological model e-Hype. Application of panel data enables to capture the regional differences. The results of the analysis include the comparison of the sensibility of different crops on climate factors in regional level.

Keywords: production function, crop yield, planted area, nonlinear VAR model, panel data.

JEL Classification: C23, C51
AMS Classification: 91B38, 91B76

1 Introduction

The analysis of agricultural production and its planning can be studied from two different points of view. The first one consists in analyzing the crop yield as a function of weather factors, the amount of used pesticides and fertilizers and the classic production factors (labor and capital). The monthly rainfall and temperature (or other climate variables like soil moisture and evaporation) could be highly significant in explaining yield but at the same time these variables are hardly predictable. The first approach helps in climate influence analysis – how much the different crops are climate sensitive in comparison with used technology of production.

The second point of view is oriented on farmers` decision-making. The main interest of farmer is to decide how many acres would be appropriate to seed. The surface of planted area could be explained by lagged variables: planted area (autoregressive model), Agricultural Producers Price Index (APPI), Consumer Price Index (CPI) and direct costs on crop production (labor, pesticides, fertilizers, fuels…). The planted area prediction model has a character of autoregressive model – the stationarity of time series must be tested and the most convenient mathematical form of model as well.

Concerning the climate sensitivity model, the influence of weather factors in the region (mainly represented by average monthly rainfall and temperature) on the real harvest will be tested. The importance of specific climate factors, i.e. evaporation and soil moisture, will be evaluated also in relation with rainfall and temperature. The crops are weather sensitive in diverse manner depending on their biological nature which is strongly related with their production function – not only the most appropriate set of variables but also the character of elasticity of substitution (unit, constant or variable).

The various mathematical forms of mentioned models will be tested in terms of their degree of conformity with reality and of course the statistical significance of the estimated parameters. Using existing panel data (Czech regions from 2002 to 2011) we will include the influence of individual differences of each region in the model, either in form of fixed or random effects.

2 Theoretical background

Production function allows to explain the output value generated by either company, industry or the whole economy based on diverse combinations of factors determining the existing technology. Detailed explanations of the Cobb-Douglas production function (PFCD) used in the analysis can be found in [4]. The same literature explains the principles of the constant elasticity of substitution production function (PFCES), including well-known

[1] University of Economics Prague, Department of Econometrics, W. Churchill Sq. 4, 13067 Prague 3, Czech
[2] Sysnet s.r.o., Kroftova 1, 15000 Prague 5, Czech Republic, bezdekovska@gmail.com.

Kmenta's approximations for two inputs. Detailed description can be found in [5]. The variable elasticity of substitution production function (PFVES) can be found in [7].

The second type of model, i.e. vector autoregression model (VAR model), is presented in [3] and the analysis of stationarity of time series is also in [2]. The farm supply theory is analyzed in [1]. Theoretical background and estimation methods for working with panel data can be found for example in [3] or [6] – fixed and random effects are sufficiently detailed. The production functions mentioned above – PFCD, PFCES and PFVES are described below.

2.1 Cobb-Douglas production function (PFCD)

The following equation shows the well-known PFCD with two production factors:

$$Y_{it} = a_i e^{gt} A_{it}^{\alpha} B_{it}^{\beta} e^{u_{it}}, \tag{1}$$

where Y_{it} is a crop yield in the region i and in the time t, a_i represents the level of achieved technology in the region i, g the non-objectified technological progress – the parameter for the proxy variable time t, A_{it} and B_{it} are the explanatory variables – production factors in the region i and in the time t. The coefficients α and β are the elasticities of output (harvest) with respect to the explanatory variables. The sum of these elasticities gives the information about the returns to scale. The u_{it} is the stochastic disturbance term. For more information about the individual effects of each region, it is useful to log-transform the model in order to obtain the form linearized in parameters. Then, we have:

$$\ln Y_{it} = \ln a_i + gt + \alpha \ln A_{it} + \beta \ln B_{it} + u_{it}. \tag{2}$$

2.2 Constant elasticity of substitution production function (PFCES)

CES production function is more general than PFCD. The elasticity of substitution may not be the unit, only needs to be constant. This function assumes returns to scale not necessarily equal to one. To capture the individual effects it is necessary to use the linearized form of the function. Given the issue and the number of relevant explanatory variables CES function for two inputs is very restrictive, but we will try to demonstrate the function application suitability using a limited set of inputs. The CES function has the following form:

$$Y = e^{gt} c \left[\gamma A^{-\rho} + (1-\gamma) B^{-\rho} \right]^{-r/\rho} e^{u}, \tag{3}$$

where e^{gt} is non-objectified technological progress, c is the parameter of efficiency of the production process, γ is the distribution parameter depending on the units of both factors A and B, r is the degree of homogeneity and ρ is the substitution parameter. The parameters can be estimated by nonlinear least squares.

As mentioned above, to incorporate the fixed or random individual effects, the linear form of the function is needed. Linear approximation is given by Kmenta in [5]. The model with individual effects has form:

$$\ln Y_{it} = \alpha_i + gt + \beta_1 + \beta_2 \ln A_{it} + \beta_3 \ln B_{it} + \beta_4 \left[\ln\left(\frac{A_{it}}{B_{it}} \right) \right]^2 + u, \tag{4}$$

where α_i is a deviation from the constant, representing the influence of each individual region i, β_1 is common constant, g the non-objectified technological progress – the parameter for the proxy variable time t, β_2 and β_3 are the elasticity coefficients of the explanatory variables A and B and β_4 expresses the elasticity of correction part of the model. The compliance of the random component with Gauss-Markov assumptions is expected. Using the estimated parameters β the estimates of the initial parameters of PFCES can be obtained:

$$c = e^{\beta_1}, \quad r = \beta_2 + \beta_3, \quad \gamma = \frac{\beta_2}{\beta_2 + \beta_3}, \quad \rho = \frac{-2\beta_4}{\beta_2 \beta_3}(\beta_2 + \beta_3). \tag{5}$$

2.3 Variable elasticity of substitution production function (PFVES)

CES production function requires the constant elasticity of substitution in all points of an isoquant sited in isoquant map. The PFVES relaxes this requirement and supposes that the elasticity of substitution is constant only along a ray drawn from zero through the isoquant map but the substitution parameter can vary along an isoquant. The VES production function has the form:

$$Y = \gamma K^{\alpha(1-\delta\rho)}\left[L + (\rho - 1)K\right]^{\alpha\delta\rho},$$
$$\gamma > 0, \quad \alpha > 0, \quad 0 < \delta < 1, \quad 0 \le \delta\rho \le 1, \quad \frac{L}{K} > \left(\frac{1-\rho}{1-\delta\rho}\right), \tag{6}$$

where Y is output, K is capital, L is labor and α, γ, δ and ρ are parameters. Detailed properties of PFVES are described in [7].

3 Model formulation and estimation

3.1 Crop weather sensitivity analysis

Regarding the aim of the article the crop production function model assumes that dependent variable, i.e. the crop yield in metric tons, is affected by different types of climate explanatory variables – the average monthly values of temperature, rainfall, evaporation and soil moisture. To estimate the production function we utilize the available annual panel data reflecting the differences in the various regions of the Czech Republic for the period between 2002 and 2011.

Table 1 shows the correlation coefficients that represent the values for the one selected region (Central Bohemia region) to illustrate the various impacts of explanatory variables on crop yield (in this case wheat). The explanatory variables in the table below correspond to the factors mentioned above (average monthly value for T3-T7 – temperature from March to July, R3-R7 – rainfall from March to July, the period between wheat sprouting and harvesting, then E6-E7 and SM6-SM7 for the evaporation and the soil moisture in June and July).

T3	T4	T5	T6	T7	R3	R4	R5	R6	R7	SM6	SM7	E6	E7
-0.24	-0.05	-0.58	-0.72	-0.40	0.14	0.29	-0.01	0.44	0.11	0.35	0.29	0.50	0.01

Table 1 Correlation coefficients between crop yield (wheat) and climate variables in Central Bohemia region

The values of correlation coefficients presented in Table 1 indicate the possibility of existence of some dependency between the crop yield and weather factors. The following subsections will focus on estimating the different crop production function based on assumptions of Cobb – Douglas, CES and VES production functions. In the following models we assume the influence of non-objectified technological progress that is expressed through a proxy variable time. The econometric software Gretl was used to estimate the parameters and statistical characteristics.

The influence of the aforementioned explanatory variables on **wheat yield** was first expressed using the PFCD. To capture the individual effects caused by the different nature of each region the form of fixed and then random effects was used. As stated above, we assume that the wheat harvest is influenced by climate factors such as average monthly temperature, rainfall, evaporation and soil moisture. Given the number of these variables and given the length of time series and the number of regions, if we would include them all at once into the model, the estimated parameter of variables that really have an important impact on the dependent variable would have not have too significant t-test or may be even insignificant. For this reason, we will examine the impact of these variables separately. The results from this partial analysis would determinate which variables are important to explain the variability of wheat yield and which are not.

Table 2 shows the OLS estimation results for temperature variables and the values reveal that all estimated parameters are significantly different from zero at 1 % confidence level except the variable designing the average monthly temperature in March that is not significant at all. On the other hand, all coefficients do not satisfy the basic assumptions of PFCD. Their values don't lie in the interval (0, 1) – they are below zero so they affect the wheat harvest negatively. Coefficient of determination and its adjusted version is very high and close to one.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
t	0.023705	4.1576	0.0001***	$\ln(a_i)$	-27.8657	-2.3919	0.0183**
$\ln(T3_{it})$	-0.023798	-0.7429	0.4590	$\ln(T6_{it})$	-0.68832	-3.4059	0.0009***
$\ln(T4_{it})$	-0.34205	-4.2013	0.0001***	$\ln(T7_{it})$	-0.70539	-11.0131	0.0000***
$\ln(T5_{it})$	-0.96309	-7.3925	0.0000***				
R-squared		0.984964		Adjusted R-squared		0.982583	

Table 2 Wheat – OLS estimation results – PFCD, fixed effects – explanatory variables t, T3-T7

The results from estimation for the PFCD with average monthly rainfall variables are similar; there is no need to state them. The evaporation and soil moisture variables were not statistically significant so we assume that those variables don't explain well the wheat yield.

As shown above, the selected production and climatic factors well explain the fluctuations in the harvest. Unfortunately, we can also conclude that the use of the PFCD to estimate wheat harvest is inappropriate, almost all of the explanatory variables violate basic assumptions of this production function. All this implies that *wheat production function has not the character of unit elasticity of substitution*. This signifies that PFCES may have the convenient attributes for estimating wheat harvest. Table 3 presents the results of OLS estimation of PFCES – the set of explanatory variables that have the most significant influence.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
t	0.02402	8.6467	0.0000***	β_1	-45.3819	-8.1041	0.0000***
$\ln(T6_{it})$	-0.56791	-5.5200	0.0000***	$\ln(T6_{it}/R7_{it})^2$	-0.03331	-2.4684	0.0150**
$\ln(R7_{it})$	0.11564	3.2364	0.0016***				
R-squared		0.587745		Adjusted R-squared		0.530299	

Table 3 Wheat – OLS estimation results – PFCES, fixed effects – explanatory variables t, T6 and R7

All estimated parameters are statistically significant at 1% confidence level except β_4 (5% confidence level). The coefficient β_4 (-0.03331) is statistically significant - if not, it would be more appropriate to use PFCD. The parameter ρ is 0.459, its positive value indicates the elasticity of substitution less than 1 and induces that translog approximation may be used. The wheat yield is explained in 59% by the model. The parameter β_2 represents the negative influence of higher temperatures in June. Contrary, the rainfall in July has a positive effect. The value of F-test (7.2351) with p-value = 2.559e-10 demonstrates the existence of important differences between regions - the model with fixed individual effects is convenient - Hausman test proved that in random effects model the parameters estimated by GLS aren't consistent. The Akaike information criterion (AIC=-253.6) has the best value from all tested models. All the results mentioned above show on *the appropriateness of the CES production function to model the wheat yield*. We apply this analyzing process to the other crops: rye, barley and corn.

Table 4 denotes results of OLS estimation of PFCD for **rye yield**– the set of explanatory variables with most significant parameters.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
$\ln(R7_{it})$	0.25016	0.11914	2.0997	0.03778**
$\ln(T3_{it})$	0.358195	0.114741	3.1218	0.00224***
β_1 – constant	6.97798	0.611034	11.4199	0.00000***
R-squared	0.809870	Adjusted R-squared		0.786870

Table 4 Rye – OLS estimation results – PFCD, fixed effects – explanatory variables T3 and R7

The estimation shows that all parameters are statistically significant. The value of F-test (38.375) with p-value = 9.88e-38 demonstrates the existence of important differences between regions. Testing the form of individual effects, Hausman test didn't reject the hypothesis that GLS estimates are consistent (for random effects). We used the Akaike information criterion to decide between two models – fixed effects (AIC=314.9) and random effects (AIC=519.1). The fixed effects are more appropriate. *All estimated parameters are the values complying the assumptions of PFCD function.* The CES and VES production functions didn't give significant results.

The results of OLS estimation of PFCD for **barley yield** are shown in Table 5.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
t	0.01362	3.2517	0.00148***	β_1	37.317	4.4776	0.00002***
$\ln(T4_{it})$	0.39928	8.7101	0.00000***	$\ln(R7_{it})$	0.08001	3.2478	0.00150***
$\mathrm{Ln}(R3_{it})$	0.043014	2.1639	0.03242**				
R-squared		0.988108		Adjusted R-squared		0.986451	

Table 5 Barley – OLS estimation results – PFCD, fixed effects – explanatory variables t, T4, R3 and R7

Table 5 shows statistical significance of all parameters. The presented estimation considered only the set of the most convenient variables. The value of F-test (756.18) with p-value=7.06e-110 demonstrates the existence of important differences between regions. The fixed individual effects showed better results (AICfixed=-185.8, AICrandom=405.8) even if Hausman test didn't reject the null hypothesis that GLS estimates in random effects model are consistent. All estimated parameters are in needed interval so the PFCD well describes the barley yield. In addition to this fact, the estimated parameters of PFCES and PFVES weren't significant.

Table 6 summarizes the results of OLS estimation for **corn yield**. As shown below, the production function of dependent variable has constant elasticity of substitution – the PFCES well describes the behavior of corn yield. Neither PFCD nor PFVES estimates were statistically significant.

Variable	Coeff.	t-Statistic	Prob.	Variable	Coeff.	t-Statistic	Prob.
β_1	14.6932	20.9719	0.00000***	$\ln(R3_{it})$	0.06478	1.8019	0.07400*
$\ln(T7_{it})$	-0.81162	-3.3196	0.00119***	$\ln(T7_{it}/R3_{it})^2$	-0.08123	-2.8572	0.0050***
R-squared		0.978812		Adjusted R-squared		0.976056	

Table 6 Corn – OLS estimation results – PFCES, fixed effects – explanatory variables T7 and R3

The high statistical significance of estimated parameters, mainly the coefficient β_4, show the appropriateness of PFCES. The corn yield is explained almost in 98%. The value of F-test (417.02) with p-value = 5.4e-95 point on the existence of important differences between regions. The type of individual effects may be evaluated by Hausman test but the test statistic can't reject the null hypothesis about consistency of GLS estimates. Random effects may be used as well as fixed. Then, we can decide between fixed and random effects using the Akaike information criterion (AICfixed = -7.02 and AICrandom = 505.59) – fixed effects are more convenient tool.

Table 7 demonstrates the **differences in weather sensitivity of different crops**. Every crop is characterized by the most important weather factors and related sensitivity coefficients that represent the intensity of influence and the direction – negative or positive.

Crop	Elasticity of substitution	R-squared	Weather factor	Sensitivity
Wheat	Constant	0.588	T6	-0.57
			R7	0.12
Rye	Unit	0.810	T3	0.36
			R7	0.25
Barley	Unit	0.988	T4	0.40
			R3	0.04
			R7	0.08
Corn	Constant	0.979	T7	-0.81
			R3	0.06

Table 7 Comparison of crop weather sensitivity

3.2 Planted area prediction model

The planted area (A) in time t could be explained by lagged variables (in time t -1): planted area (autoregressive model), Agricultural Producers Price Index (APPI), Consumer Price Index (CPI) and direct costs (DC) on crop production (labor, pesticides, fertilizers, fuel…). The estimation method for VAR model presumes the stationarity of time series. The non-stationarity may be eliminated by including time variable or first differences of dependent variable. The planted area doesn't evince any visible trend. We applied the unit root test (Augmented

Dickey-Fuller test) that rejected the null hypothesis that all groups have unit root. The VAR model form of linear regression and power-law function applied on panel data of wheat crop was tested. The non-linear function (Cobb-Douglas type) showed better results. The linearized model has a form:

$$\ln A_{it} = gt + c_1 + c_2 \ln A_{i,t-1} + c_3 \ln APPI^1_{i,t-1} + c_4 \ln APPI^2_{i,t-1} + c_5 \ln CPI_{i,t-1} + c_6 \ln DC_{i,t-1} + u_{it}, \quad (7)$$

where t is time variable and g its parameter, c_1 is a constant and parameters c_2 to c_6 correspond to the influence of lagged dependent (A) and explanatory variables: price index of a studied crop - wheat $APPI^1$, price index of a competing crop – rye $APPI^2$, CPI and direct costs DC. Table 8 presents the results of OLS estimation.

Variable	Coefficient	t-Statistic	Prob.	Variable	Coefficient	t-Statistic	Prob.
g	0.02046	8.2220	0.00000***	c_4	-0.12009	-4.3238	0.00003***
c_1	-27.9551	4.7445	0.00000***	c_5	1.14872	5.4983	0.00000***
c_2	-0.18314	-2.9331	0.00411***	c_6	-0.47630	-7.6641	0.00000***
c_3	-0.13410	-4.8649	0.00000***				
R-squared		0.996628		Adjusted R-squared		0.996024	

Table 8 Wheat – OLS estimation results – Planted area prediction model, fixed effects

As shown above, the estimated parameters are all statistically significant at 1% confidence level. The model well describes the variability of dependent variable. The high value of determination coefficient R-squared may denote the spurious regression. Granger and Newbold [2] show that the relation R-squared > DW statistic may signify that the residuals have non-stationary character. In that case, DW statistic is 1.202 and mentioned hypothesis could be rejected. In evaluation of the importance of regional individual effects, the F – test (F = 27.64, p-value = 3.22e-28) rejected the null hypothesis that the groups have common intercept – the individual effects of different regions have important influence. The model with random effects doesn't explain the variability of dependent variable because the Hausman test rejected the null hypothesis that GLS estimates are consistent. The Akaike information criterion (AIC = -330.26) for the fixed effects model has the best value in comparison with other tested forms of model. Due to the use of panel data, low value of DW statistic does not necessarily mean that the autocorrelation is present.

4 Conclusion

The paper presents the analysis of crop weather sensitivity and planted area prediction model. The first part of paper demonstrates that every crop needs to be considered separately. Wheat and corn could be represented by PFCES but the production of rye and barley has character of unit elasticity of substitution – PFCD. Every analyzed crop was influenced by different weather factors and their importance varies – due to the crop biological nature. The second part of the paper concentrates on planted area predictions. The tests show that the model has autoregressive character and has the form of power-law function. The important influence of lagged dependent and explanatory variables (agricultural producers price index for the studied crop and for the competing crop, consumer price index and direct cost of production) is proved.

Acknowledgements

This work was supported by Internal Grant Agency of the University of Economics in Prague - IGA F4/18/2012 and project SUDPLAN (European Union).

References

[1] Ferris, J.N.: *Agricultural prices and commodity market analysis*. Michigan State University Press, East Lansing, Michigan, 2005.
[2] Granger, C.W., J.-Newbold, P.: Spurious Regression in Econometrics, *Journal of Econometrics* **2** (1974), 111-120.
[3] Greene, W.H.: *Econometric analysis, 5th ed.* Prentice Hall, Upper Saddle River, New York, 2003.
[4] Hušek, R.: *Aplikovaná ekonometrie.* Vysoká škola ekonomická v Praze, Praha, 2001.
[5] Kmenta, J.: On estimation of the CES production Function. *International Economic Review* **8** (1967),80-89.
[6] Koop, G.: *Introduction to Econometrics.* John Wiley & Sons Ltd., England, 2008.
[7] Revankar, N.S.: A Class of Variable Elasticity of Substitution Production Functions. *Econometrica* **39** (1971), 61-71.