

Problem of heteroskedasticity in econometric models of stocks in farms

Boleslaw Borkowski*, Hanna Dudek*

Introduction

In free-market economy level of stocks in farms is a very important problem. It has affected efficiency at managing by charges of their storage cost on the one hand and safety of farm activity on the other hand. Optimal level of stocks depends on many factors. Generally, stocks in actual period depend on economical activity in this period and on stocks in previous period.

Desired level of inventories can be computed from econometric model (Weife, Ramanathan, Green):

$$R_t^* = b_0 + b_1 ATK_t + b_2 R_{t-1} + \xi_t$$

where:

R_t^* - desired level of inventories in actual period,

ATK_t - level of economical activity in actual period,

R_{t-1} - level of inventories in previous period,

ξ_t - error term,

$b_1 > 0$ and $b_2 > 0$, $b_2 < 1$.

Required stocks in farms are considered taking into account the place of their appearing and destination. Usually level of inventory is expressed in the natural units, therefore models for each category of stocks in farms (grain, potatoes, hay, silage etc.) should be considered separately.

The main interest of managing was put here on the cope of the stocks in question per farm. The methodological problem of estimation of econometric models is tightly connected with the matter. Having in mind that 60% of arable land is designed for corn's cultivate one can expect that there should be a great difference in the area of corn's cultivating as well as in the scope of stocks. According to the researches, the level of stock of corn grains is directly propor-

* Warsaw University.

tional to the area of arable land. Distribution of area of corn cultivating has been right handly asymmetric what meant that the biggest area of corn's cultivating was situated in largest farms. Bigger farms usually can storage more stocks of agricultural products. It effects that in econometric models of storing, error terms corresponding to large farms have large variances. Lack of constancy of variances (known as heteroskedasticity) commonly can be met at cross-section data. If heteroskedasticity among the stochastic error terms in econometric model is ignored and ordinary least squares procedure is used to estimate the parameters, then the estimators based on it will still be unbiased and consistent but inefficient. Moreover the estimated variances and covariances of the regression coefficients will be biased and incontistent. and hence a lot of common tests of hypotheses are invalid. In such cases generalized least squares procedure is recommended.

The problem of heteroskedasticity has been explained upon the example of grain's storing.

Data

Inventory data was collected in frame of KBN (State Committee for Scientific Researches) grant No 5 P06J01117 "Managing of stocks in farms". Data used for building model were taken from 1998 and 1999 years from central-western macro-region in Poland. There have been taken under consideration 80 farms of 15 up to 55 ha. the arable land in this study. The concerned farms kept accountancy in co-operation with Institute of Economics of Agriculture and Food Economy. Used sample is not a representative one. Obtaining of random size was not available due to lack of precise list of indeed existing individual farms in Poland and the principle free participation farmers in researches. The main direction of choosing the farms was their typical structure of area dominating in that macro-region.

Estimation procedure

The multiple linear regression model has the following general formulation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon ,$$

where

Y denotes dependent variable,

X_j - denotes independent variable, $j= 1,2,\dots,k$,

β_j are unknown parameters to be estimated, $j=0, 1,\dots,k$,

ε is unobserved error term.

In ordinary least squares procedure vector of estimated regression coefficients b is given by

$$b = (X^T X)^{-1} X^T y,$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ . & . & . & \dots & . \\ . & . & . & \dots & . \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{bmatrix}, b = \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ b_k \end{bmatrix}$$

x_{ji} denotes i th observation of X_j , $j=1, 2, \dots, k$,

y_i denotes i th observation of Y , $i=1, 2, \dots, n$,

n - sample size,

X is nonsingular matrix of observations,

X^T is transpose of matrix X ,

$(X^T X)^{-1}$ denotes inverse matrix of $X^T X$.

In ordinary least squares procedure covariance matrix of the estimator b is given by:

$$D^2(b) = S_e^2 (X^T X)^{-1},$$

where

$$S_e = \frac{e^T e}{n-k-1} \text{ - standard error of the estimation,}$$

e - vector of residuals.

Under appropriate assumptions (Green, Judge, Ramanathan) ordinary least squares estimators are most efficient among unbiased linear estimators. One of important assumption refers to variance of error term all ε 's should be identically distributed with the same variance σ^2 . This is known as homoskedasticity. In many situations commonly encountered with cross-section data, however, this assumption might be false. If heteroskedasticity among the stochastic error terms in econometrical model is ignored and ordinary least squares procedure is used to estimate the parameters, then the estimators based on it will still be unbiased and consistent but inefficient. Moreover the estimated variances and covariances of the regression coefficients will be biased and inconsistent, and hence a lot of common tests of hypotheses are invalid. In such cases generalized least squares procedure is recommended.

If all assumptions of ordinary least squares method are fulfilled, covariance

matrix is given by:

$$\begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & & \\ \dots & & \dots & \\ 0 & & & \sigma^2 \end{bmatrix}$$

In the case of heteroskedasticity, assuming that error terms are pairwise uncorrelated.

Covariance matrix can be written as

$$\begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \\ \dots & & \dots & \\ 0 & & & \sigma_n^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & 0 & 0 \\ 0 & \omega_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \omega_n \end{bmatrix} = \sigma^2 \Omega$$

where

$$\sigma_i^2 = \sigma^2 \omega_i \quad i=1, 2, \dots, n.$$

The generalized least squares estimator is given by

$$a = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y,$$

and its covariance matrix takes form:

$$D^2(a) = S^2 (X^T \Omega^{-1} X)^{-1},$$

Where

$$S^2 = \frac{e^T \Omega^{-1} e}{n - k - 1}$$

The generalized least squares estimates obtained in this way are consistent and so are the estimated variances and covariances of the estimates. The estimates are asymptotically likely to be more efficient than ordinary least squares estimates.

The generalized least squares procedure applied to the case of heteroskedasticity is also known as weighted least squares procedure. The weights are defined as reciprocals of standard deviation of error. Observations with a relatively low standard deviation of error term σ_i are more reliable, are weighted more heavily, and hence play a greater role in the estimation process than those that are less reliable because they have relatively high σ_i .

In practice matrix $\sigma^2 \Omega$ is unknown. A researcher must first obtain estimates of σ_i , divide each observation by it and then use the ordinary least squares procedure to the transformed observations.

Empirical results

In the model of storing of grain the dependent variable is stock of grain in 1999. By using stepwise regression procedure in Statgraphics statistical package from 20 potential variables we have selected 3 independent variables.

Using ordinary least squares procedure we obtained the following model:

$$Y_t = -31138 + 0,506Y_{t-1} + 192,91X_{1t} + 1286,8X_{2t} + 584,07X_{3t}$$

(5216,2) (0,075) (93.1) (242,4) (146,56)

where

Y_t - stock of grain in 1999, [kg.]

Y_{t-1} - stock of grain in 1998, [kg.]

X_{1t} - number of live-stock in 1999 [big head],

X_{2t} - area of corns in 1999. [ha.]

X_{3t} - crop of corns in 1999, [dt/ha.]

values in parentheses are standard errors of regression coefficients.

In this model two outliers were eliminated.

Table 1 Results for model estimated ordinary least squares procedure

Symbol of variable	t statistic	p-value	partial R^2
intercept	- 5,969	0,0000	0,3280
Y_{t-1}	6,742	0,0000	0,3838
X_{1t}	2,074	0,0416	0,0557
X_{2t}	5,309	0,0000	0,2785
X_{3t}	3,985	0,0002	0,1787

Source: Own computing by using Pc-Give statistical package.

All parameters are significantly different from zero at the 5 percent level of significance.

Coefficient of determination $R^2 = 0,83$.

Estimated model explains in 83% the total variation of corn stocks. The greatest importance for the explanation of variability of dependent variable had the stocks from the previous year.

They have got a great influence on the level of in the year in question (average of about 0,5 kg for each kg of increasing of stocks from the previous year). The increase in stock of 1 big head affected in picking up the level of stocks of grain by 193 kg average.

In ordered sample considering area of the arable land, Durbin-Watson statistic $DW = 1,81$. Critical values for the 5 percent level of significance and number of independent variables $k = 4$ and sample size $n = 78$ equals: $DW_1 = 1,52$ and $DW_u = 1,74$, then there is probably not any serious correlation.

Statistical tests proved that error term is a normally distributed random variable. Shapiro-Wilk statistic equals 0,9920, (at the 5 percent level of significance critical value 0,9597, p-value = 0,9868) and Jarque-Bera statistic is 4,5385 at critical value 5,9910 (p-value = 0,1034).

Heteroskedasticity was tested by using Goldfeld – Quandt, White and Harvey – Godfrey tests.

- In Goldfeld-Quandt test we have divided the sample into first 25 and last 25 observations. We have computed $F = 4,9137$, so we have rejected the null hypothesis of homoskedasticity and conclude that heteroskedasticity is present (critical values for the 5 percent level of significance and $m_1 = 20$, $m_2 = 20$ degrees of freedom $F_{\alpha} = 2,12$).
- In White test computed $\chi^2 = 58,385$ at critical values for the 5 percent level of significance and 14 degrees of freedom $\chi^2_{\alpha} = 23,685$.
- In Harvey - Godfrey test obtained $\chi^2 = 25,436$ at critical values for the 5 percent level of significance and 2 degrees of freedom $\chi^2_{\alpha} = 5,991$.

Upon all these tests we have found presence of heteroskedasticity.

Then we have applied generalized least squares procedure. The formal steps were as follows.

1. From model estimated by using ordinary least squares procedure we have calculated the residuals e_1, e_2, \dots, e_n .
2. We have regressed logarithms of these residuals against exogenous variables:

X_3 – crop of corns in 1999, [dt/ha.]

X_4 – area of arable land in 1999, [ha.]

$$\ln(e^2) = -323,06 + 6,50 X_3 + 8,16 X_4$$

$$(79,76) \quad (2,09) \quad (2,11),$$

values in parentheses are standard errors of regression coefficients.

This is the auxiliary regression model, in which independent variables were selected by using stepwise regression procedure.

3. We have taken antilog to get predicted variances.
4. We have estimated model by weighted least squares procedure. The weights are defined as reciprocals of predicted standard deviation of error.

Model estimated by using generalized least squares procedure has form:

$$Y_t = -27114,3 + 0,519 Y_{t-1} + 178.365 Y_{1t} + 1194,8 Y_{2t} + 497,606 Y_{3t}$$

(311967) (00426) (48,0681) (142,099) (77,1373)

Coefficient of determination $R^2 = 0,90$ (R^2 was computed as square of the correlation between observed and predicted values of dependent variable).

Durbin-Watson statistic $DW = 1,82$. Shapiro-Wilk statistic equals 0,9763, (a, the 5 percent level of significance critical value 0,9597, p-value = 0,4334). Statistical tests proved that error term is a normally distributed random variable and there is probably not any serious correlation.

Heteroskedasticity was tested by using Goldfeld - Quandt, White and Harvey - Godfrey tests.

- In Goldfeld-Quandt test we have divided the sample into first 25 and last 25 observations. We have computed $F = 2,09$ (critical values for the 5 percent level of significance and $m_1 = 20$, $m_2 = 20$ degrees of freedom $F_{\alpha} = 2,12$).
- In White test computed $\chi^2 = 17,74$ at critical values for the 5 percent level of significance and 14 degrees of freedom $\chi^2_{\alpha} = 23,685$.
- In Harvey - Godfrey test obtained $\chi^2 = 4,33$ at critical values for the 5 percent level of significance and 2 degrees of freedom $\chi^2_{\alpha} = 5,991$.

In this case we have carried statistical test and we have found normality and homoskedasticity.

The estimated model meets all requirements for a quality model.

Standard errors of the regression coefficients were lower than those obtained by the least squares procedure. Applying of generalized least squares procedure for estimation of parameters of model improved its quality.

Upon obtained results we can state that the increase by 1 leg of stocks of corns from the previous year (holding other variables constant) affected increasing average of about 0,5 kg of stocks in the year in question.

The number of the live-stock in the farms had a big influence for the level of stocks of corns.

The increase in stock of 1 big head affected in picking up the level of stocks of grain by 178 kg average.

The impact of area of corns and crops of corns was average about 1195 and 498 adequately.

Our researches have shown that level of storage of agricultural products in farms depends on level of storage of this products in previous period and level of economical activity (number of live-stock, area of corns, crops of corns).

Conclusions

Our researches have shown that in econometric models of storing in farms variables can be used per-farm, not per-hectare terms. Heteroskedasticity can be eliminated in such models by using for the estimation of parameters generalized least squares procedure.

As the structure of the heteroskedasticity is generally unknown, a researcher must first obtain estimates of σ_i and then use the weighted least squares procedure. The natural method of estimation the residual standard deviations is to exploit the information contained in the auxiliary regression. The main problem refers to selection of explanatory variables in this model. In agricultural economics researches one should suppose that in set of these variables is area of arable land or variable tightly correlated with it.

References

- ACZEL A. D., *Complete Business Statistics*, Irwin, Boston, 1993.
- BERNDT E. R., *The Practice of Econometrics: Classic and Contemporary*, Addison-Wesley Publishing Company, New York, 1996.
- GREENE W. H., *Econometric Analysis*, Prentice Hall, New York, 2000.
- JUDGE G. G., HILL C., GRIFFITHS W. E., LUTKEPOHL H., LEE T., *Introduction to the Theory and Practice of Econometrics*, John Wiley&Sons, New York, 1988.
- RAMANATHAN R., *Introductory Econometrics with Applications*, The Dryden Press, Fort Worth, 1998.
- SILVER E. A., PETERSON R., *Decision Systems for Inventory Management and Production Planning*, John Wiley&Sons, New York, 1985.
- THEIL H., *Principles of Econometrics*, John Wiley&Sons, New York, 1971.
- WELFE W., WELFE A., *Applied Econometrics* (in Polish), PWE, Warsaw, 1996.