Discussiones Mathematicae
Probability and Statistics 29 (2009) 41–51

ON INCONSISTENCY OF HELLWIG'S VARIABLE CHOICE METHOD IN REGRESSION MODELS

Tadeusz Bednarski and Filip Borowicz

Institute of Economic Sciences, Faculty of Law,
Administration and Economics, University of Wroclaw
Uniwersytecka 22/26, 50–145 Wrocław

e-mail: t.bednarski@prawo.uni.wroc.pl
e-mail: f.borowicz@prawo.uni.wroc.pl

Abstract

It is shown that a popular variable choice method of Hellwig, which is recommended in the Polish econometric textbooks does not enjoy a very basic consistency property. It means in particular that the method may lead to rejection of significant variables in econometric modeling. A simulation study and a real data analysis case are given to support theoretical results.

Keywords and phrases: model choice, econometric modeling.

2000 Mathematics Subject Classification: Primary 62F99; Secondary 62P20.

1. Introduction

Model selection methods have shown to be very useful in applications of regression models. They are of special interest in econometric modeling where – under relatively small sample sizes – larger sets of explanatory variables are rule rather than exception. Such methods of variable selection like Akaike or Schwarz are today standards in statistical analysis and they are available in commercial statistical packages. The methods satisfy basic asymptotic consistency criteria and are even proved to be optimal – like the method of Schwarz (1978). It is shown here that a popular variable choice method of Hellwig which is the most frequently recommended in the Polish econometric textbooks may not enjoy this basic consistency property in very elementary situations.

Assume that parameters indexing probability distributions of a statistical regression model are elements of R^d and suppose we observe a sample of size n. Let $\Theta^1, \Theta^2, \ldots, \Theta^J$ be all different sub-models of regression type – for simplicity we suppose that these are linear subspaces of R^d. Denote by $d(j)$ the dimension of Θ^J and by $\rho_F(\theta)$ some objective function – aimed to measure a distance between the underlying and the model distribution – which depends both on the true distribution F and on the regression parameter θ. Let ρ_F^j be the minimum value of $\rho_F()$ restricted to the sub-model Θ^j. The model distribution F may be replaced by the corresponding empirical distribution F_n, which is based on a sample of size n. The Akaike (1969) model selection method (AIC) chooses the model Θ^j for which $S_n(j) = n\rho_{F_n}^{(j)} + d(j)$ takes the minimum value, where $n\rho_{F_n}^{(j)} = -\log L(j, \hat{\theta}^j)$ and $L(j, \hat{\theta}^j)$ is the likelihood function corresponding to j-th model with parameter θ replaced by its maximum likelihood estimator. For the Schwarz (1978) criterion $S_n(j) = n\rho_{F_n}^{(j)} + 0.5d(j)\log n$ is used. The two methods enjoy natural conditions of asymptotic consistency of variable selection – conditions commonly recognized as the "minimal". Namely, if j^* denotes the index of the correct model of minimal dimension then for all $j \neq j^*$ the probability $P(\{S_n(j) - S_n(j^*) > 0\})$ converges to 1 as sample size n tends to infinity and j corresponds to incorrect model (see Machado (1993) and Bednarski & Mocarska (2006) for general conditions leading to asymptotic consistency of model selection methods).

Hellwig (1969) proposed a variable choice method for the classical linear regression model based on the following criterion: choose those explanatory variables in the linear regression model $Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_kX_k + \varepsilon$ for which the expression

$$\sum_{i \in H} \frac{\rho^2(Y, X_i)}{1 + \sum\limits_{\substack{m,i \in H \\ m \neq i}} |\rho(X_i, X_m)|}$$

takes its maximum value, where maximisation is over all possible subsets of indexes of explanatory variables H and ρ denotes the correlation coefficient.

The empirical implementation of the above formula involves the empirical correlation coefficients. The criterion has a very intuitive appeal since it tends to select a set of explanatory variables weakly correlated among themselves and highly correlated with the dependent variable. It was shown however that the approach may be misleading in time series analysis (Serwa (2004)). In the following section we demonstrate that the method may in fact fail in much simpler situations. Section 2 shows the inconsistency of Hellwig's method. Section 3 demonstrates results of a simulation experiment comparing the efficiency of the basic model choice methods with Hellwig's method. We also include a real econometric data analysis case.

2. Inconsistency of Hellwig's method

Since Hellwig's method is the first place model selection method recommended in the Polish academic econometric literature it is likely that the method may also be used in real data studies. We show here that Hellwig's criterion has a crucial drawback, it need not lead to asymptotic consistency as defined in the previous section – a wrong sub-model can be selected under some model conditions.

Let us define the empirical objective function corresponding to Hellwig's method as

$$S_n(H) = -\sum_{i\in H} \frac{\rho_n(Y, X_i)}{1 + \sum\limits_{\substack{m,i\in H\\ m\neq i}} |\rho_n(X_i, X_m)|},$$

where H is a regression sub-model or equivalently a subset of explanatory variables and ρ_n is the correlation coefficient for two selected variables corresponding to a sample of size n. By the sample we mean here independent and identically distributed random vectors $(Y_1, X_{11}, \ldots, X_{k1})$, $\ldots$, $(Y_n, X_{1n}, \ldots, Y_{kn})$. Consider also the following simple linear regression model $Y = aX_1 + bX_2 + \varepsilon$, where a, b are non zero structural parameters, $X_1 = X + \varepsilon_1$, $X_2 = X$ while the variables X, ε, ε_1 have positive variances and are independent with expectations equal zero. Obviously the minimum dimension model H^* here is the full dimension model with variables X_1, X_2.

Fact. Under the above model conditions we can always find positive a, b such that the probability $P(\{S_n(H) - S_n(H^*) < 0\})$ will converge to one, where H contains only X_1.

Proof. Since empirical correlations converge with probability one to population correlations we shall compare the quantities

$$S_n(H) = -\sum_{i \in H} \frac{\rho^2(Y, X_i)}{1 + \sum\limits_{\substack{m,i \in H \\ m \neq i}} |\rho(X_i, X_m)|}.$$

To prove that $P(\{S_n(H) - S_n(H^*) < 0\})$ converges to one it will suffice to show that for some a and b

$$\frac{\rho^2(Y, X_1) + \rho^2(Y, X_2)}{1 + |\rho(X_1, X_2)|} < \rho^2(Y, X_1).$$

A simple calculation leads to the following formulas

$$\rho(X_1, X_2) = \frac{\sigma_x}{\sqrt{\sigma_x^2 + \sigma_1^2}},$$

$$\rho(Y, X_1) = \frac{(a+b)\sigma_x^2 + a\sigma_1^2}{\sqrt{\sigma_x^2 + \sigma_1^2} \cdot \sqrt{(a+b)^2\sigma_x^2 + a^2\sigma_1^2 + \sigma_\varepsilon^2}},$$

$$\rho(Y, X_2) = \frac{(a+b)\sigma_x^2}{\sqrt{\sigma_x^2} \cdot \sqrt{(a+b)^2\sigma_x^2 + a^2\sigma_1^2 + \sigma_\varepsilon^2}},$$

where σ_x^2, σ_1^2, σ_ε^2 are variances of X, ε_1, ε.

Now notice that

$$\frac{\rho^2(Y, X_1) + \rho^2(Y, X_2)}{1 + |\rho(X_1, X_2)|} < \rho^2(Y, X_1)$$

is equivalent to

$$\frac{\rho^2(Y, X_2)}{\rho^2(Y, X_1)} < |\rho(X_1, X_2)|.$$

Plugging in the above correlations into the last inequality we obtain

$$\left[\frac{(a+b)\sigma_x^2}{(a+b)\sigma_x^2 + a\sigma_1^2}\sqrt{\frac{\sigma_x^2+\sigma_1^2}{\sigma_x^2}}\right]^2 < \sqrt{\frac{\sigma_x^2}{\sigma_x^2+\sigma_1^2}}$$

and consequently

$$(1) \qquad \left[\frac{\sigma_x^2}{\sigma_x^2 + \dfrac{a}{a+b}\sigma_1^2}\right]^2 < \left[\frac{\sigma_x^2}{\sigma_x^2+\sigma_1^2}\right]^{3/2}.$$

Since $\frac{\sigma_x^2}{\sigma_x^2+\sigma_1^2} < 1$ the above inequality holds if for instance $\frac{a}{a+b} > 1$. This completes the proof.

Remark 1. A similar reasoning gives inconsistency of the method when we have a negative correlation between the variables X_1, X_2, more precisely when $X_1 = -X + \varepsilon_1$ and $X_2 = X$. Then in the above formulas we would have to change the expression $(a + b)$ into $(b - a)$. Notice that the requirement $\frac{a}{a+b} > 1$ covers the range of situations that are by no means marginal in practical modeling. Moreover since the value $\rho(X_1, X_2) = \frac{\sigma_x}{\sqrt{\sigma_x^2+\sigma_1^2}}$ is unrelated to $\frac{a}{a+b} > 1$ the regression models given above need not be affected by collinearity to lead to a wrong model choice.

Remark 2. It is shown below that adding a set of mutually uncorrelated explanatory variables to X_1, X_2 does not improve the efficiency of Hellwig's method here. The expression

$$\frac{\rho^2(Y,X_1)+\rho^2(Y,X_2)}{1+|\rho(X_1,X_2)|}<\rho^2(Y,X_1)$$

changes then to

$$\frac{\rho^2(Y,X_1)+\rho^2(Y,X_2)+\ldots+\rho^2(Y,X_k)}{1+|\rho(X_1,X_2)|}$$

$$<\rho^2(Y,X_1)+\rho^2(Y,X_3)+\ldots+\rho^2(Y,X_k)$$

and it simplifies to

$$\frac{\rho^2(Y,X_2)}{\rho^2(Y,X_1)+\rho^2(Y,X_3)+\ldots+\rho^2(Y,X_k)}<|\rho(X_1,X_2)|.$$

Plugging in the correlations into the above formula gives

$$\frac{\left(\dfrac{(a+b)\sigma_x^2}{\sqrt{\sigma_x^2}\cdot\sqrt{(a+b)^2\sigma_x^2+a^2\sigma_1^2+\sigma_\varepsilon^2}}\right)^2}{\left(\dfrac{(a+b)\sigma_x^2+a\sigma_1^2}{\sqrt{\sigma_x^2+\sigma_1^2}\cdot\sqrt{(a+b)^2\sigma_x^2+a^2\sigma_1^2+\sigma_\varepsilon^2}}\right)^2+\rho^2(Y,X_3)+\ldots+\rho^2(Y,X_k)}$$

$$<\frac{\sigma_x}{\sqrt{\sigma_x^2+\sigma_1^2}}$$

and finally

$$\frac{\left[\dfrac{\sigma_x^2}{\sigma_x^2+\frac{a}{a+b}\sigma_1^2}\right]^2}{1+\dfrac{\left((a+b)^2\sigma_x^2+a^2\sigma_1^2+\sigma_\varepsilon^2\right)\left(\sigma_x^2+\sigma_1^2\right)\left(\rho^2(Y,X_3)+\ldots+\rho^2(Y,X_k)\right)}{\left((a+b)\sigma_x^2+a\sigma_1^2\right)^2}}$$

$$<\left[\frac{\sigma_x^2}{\sigma_x^2+\sigma_1^2}\right]^{3/2},$$

which is always satisfied if only (1) holds. The variable X_2 will then be removed from the full model as nonsignificant if only $\frac{a}{a+b}>1$.

3. Simulation study and real data analysis

A simulation experiment was carried out to compare Hellwig's method with Akaike and Schwartz's procedures. Samples of size 10, 100 and 500 from the model $Y=aX_1+bX_2+\varepsilon$ were considered. Explanatory variables were either independent or dependent. For each model case and sample size the selection procedure was repeated 10000 times.

In the dependent case the explanatory variables $X_1=X+\varepsilon_1$ and $X_2=X$ where X, ε, ε_1 are independent standard normal variables. Frequencies of correct variable selection for the model $Y=X_1-0.5X_2+\varepsilon$ are presented in Table 1. Table 2 gives results in the case $Y=X_1+0.5X_2+\varepsilon$. Apparently Hellwig's method is very inferior as compared to Akaike's and Schwartz's procedure.

Table 1. Frequency of correct variable selection for the model

$$Y=X_1-0.5X_2+\varepsilon.$$

sample size	10	100	500
Hellwig's method	0.0924	0	0
Akaike's method	0.4266	0.9779	1
Schwartz method	0.3891	0.9044	1

Table 2. Frequency of correct variable selection for the model

$$Y = X_1 + 0.5X_2 + \varepsilon.$$

sample size	10	100	500
Hellwig's method	0.4823	0.5705	0.6496
Akaike's method	0.3638	0.9796	1
Schwartz method	0.3216	0.9051	1

Table 3 shows results for independent covariates – X_1, X_2, ε are independent standard normal. In this case frequencies of correct variable selection for the model $Y = X_1 + 0.5X_2 + \varepsilon$ are roughly similar for all the methods.

Table 3. Frequency of correct variable selection for the model

$$Y = X_1 + 0.5X_2 + \varepsilon.$$

sample size	10	100	500
Hellwig's method	0.5271	0.9006	0.9983
Akaike's method	0.5112	0.9995	1
Schwartz method	0.4713	0.9952	1

In addition a real data set from OECD (source "Economic Outlook No 84: Annual and Quarterly data") was also used to compare Hellwig's and Akaike's selection efficiency. The following two models were analyzed:

$$GDP = a \cdot Import + b \cdot Export + \varepsilon,$$

$$GDP = a \cdot EmployGov + b \cdot EmployTotal + c \cdot UnEmploy + \varepsilon,$$

where variable GDP is gross domestic product (volume, at 2000 PPP, USD), $Import$ is imports of goods and services (value, USD), $Export$ is exports of goods and services (value, USD), *EmployGov* is general government employment, *EmployTotal* is total employment and *UnEmploy* is unemployment rate. The analysis of the first model was based on data from 2006 from all 30 OECD member countries. In the second case not all the data were available for all 30 countries, so analysis was reduced to 24 OECD countries. Tables 4 and 5 show results of this analysis. Data from 2005 to 2000 ware also analyzed and results were very similar.

Table 4. Result of the analysis of the model

$$GDP = a \cdot Import + b \cdot Export + \varepsilon.$$

method	selected model
Hellwig's	$GDP = b\, Export + \varepsilon$
Akaike's	$GDP = a\, Import + b\, Export + \varepsilon$

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.991e+04  1.228e+05  -0.813    0.423
Import       8.720e+00  6.911e-01  12.618 7.80e-13
Export      -5.634e+00  8.652e-01  -6.511 5.56e-07

Residual standard error: 477500 on 27 degrees of freedom
Multiple R-squared: 0.9516,     Adjusted R-squared: 0.948
F-statistic: 265.2 on 2 and 27 DF,  p-value: < 2.2e-16

Correlations:
             GDP    Import    Export
GDP    1.0000000 0.9356814 0.8160590
Import 0.9356814 1.0000000 0.9573338
Export 0.8160590 0.9573338 1.0000000
```

Table 5. Result of the analysis of the model

$$GDP = a \cdot EmployGov + b \cdot EmployTotal + c \cdot UnEmploy + \varepsilon.$$

method	selected model
Hellwig's	$GDP = a\, EmployGov + b\, EmployTotal + c\, UnEmploy + \varepsilon$
Akaike's	$GDP = b\, EmployTotal + c\, UnEmploy + \varepsilon$

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.780e+05  2.130e+05   1.305 0.206625
EmployGov   1.155e-03  2.537e-02   0.046 0.964147
EmployTotal 1.394e+00  3.051e-01   4.568 0.000187
UnEmploy    7.958e-01  2.897e-01   2.747 0.012431

Residual standard error: 725700 on 20 degrees of freedom
Multiple R-squared: 0.9768,     Adjusted R-squared: 0.9734
F-statistic: 281.2 on 3 and 20 DF,  p-value: < 2.2e-16

Correlations:                      GDP EmployGov EmployTotal
UnEmploy
GDP         1.0000000 0.9698061   0.9827763 0.9483883
EmployGov   0.9698061 1.0000000   0.9763534 0.9387297
EmployTotal 0.9827763 0.9763534   1.0000000 0.9242938
UnEmploy    0.9483883 0.9387297   0.9242938 1.0000000
```

Notice that in the case of the first model, even though variables *Import* and *Export* are very highly correlated, Akaike's method logically retians both of the them since both are significant. In the second case however Akaike's method rejects the variable *EmployGov* since it is not significant in the full model and moreover it is highly correlated with total employment.

References

[1] H. Akaike, *Information theory and an extension of the maximum likelihood principle*, In Second International Symposium on Information Theory, Budapest: Academiai Kiado (1973), 267–81.

[2] T. Bednarski and E. Mocarska, *On robust model selection within the Cox model*, Econometrics Journal **9**, (2006), 279–290.

[3] Z. Hellwig, *Problem optymalnego doboru predyktant*, Przegląd Statystyczny nr 4 (1969).

[4] J.A.F. Machado, *Robust model selection and M-estimation*, Econometric Theory **9** (1993), 478–493.

[5] G. Schwarz, *Estimating the dimension of a model*, The Annals of Statistics **6** (2) (1978), 461–64.

[6] D. Serwa, *Metoda Hellwiga jako kryterium doboru zmiennych do modeli szeregów czasowych*, Szkoła Główna Handlowa, Kolegium analiz Ekonomicznych, Instytut Ekonometrii 2004.

Received 15 March 2009