CENTRAL BANK OF CYPRUS
EUROSYSTEM

WORKING PAPER SERIES

Mean Squared Prediction Error Reduction with Instrumental Variables

Antonis Michis

July 2016

Working Paper 2016-05

Address
80 Kennedy Avenue
CY-1076 Nicosia, Cyprus

Postal Address
P. O. Box 25529
CY-1395 Nicosia, Cyprus

E-mail
publications@centralbank.gov.cy

Website
http://www.centralbank.gov.cy

Fax
+357 22 378153

Papers in the Working Paper Series may be downloaded from:
http://www.centralbank.gov.cy/nqcontent.cfm?a_id=5755

MEAN SQUARED PREDICTION ERROR REDUCTION WITH INSTRUMENTAL VARIABLES

Antonis A. Michis*

Abstract

The mean squared prediction error of the linear regression model is examined when estimation is performed with instrumental variables. It is shown that increasing the number of instruments in the estimation procedure, can reduce the mean squared prediction error of the model through more efficient estimation of the coefficient vector.

Keywords: Mean squared prediction error; efficiency; instrumental variables

JEL classification: C01; C26; C53

---

* Statistics Department, Central Bank of Cyprus, 80 Kennedy Avenue, CY-1076, Nicosia, Cyprus. Email: AntonisMichis@centralbank.gov.cy, https://sites.google.com/site/antonismichis/ .

The author would like to thank Professor Franco Peracchi for helpful comments. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the Central Bank of Cyprus or the Eurosystem.

1. Introduction

Consider the linear regression model

$$y = X\beta + \varepsilon\,, \qquad E(\varepsilon\varepsilon^{'}) = \Omega \tag{1}$$

where y is a $T \times 1$ vector of time observations, X is a $T \times K_1$ matrix of observations on K_1 explanatory variables, ε is the $T \times 1$ vector of disturbances which satisfies the orthogonality condition and Ω is the error covariance matrix. Assume also a $T \times K$ matrix $Z = [Z_X, Z_Z]$ of instrumental variables exists. It includes both the explanatory variables in $X = Z_X$ and the additional instruments Z_z derived from deterministic functions of them (see for example Cragg, 1983) that satisfy the necessary identifiability conditions (Wooldridge, 2010 p. 99).

Consider also the generalized two stage least squares estimator (see, Johnston and DiNardo, 1997, p. 337)

$$\hat{\beta} = (X^{'}Z\Sigma^{-1}Z^{'}X)^{-1}X^{'}Z\Sigma^{-1}Z^{'}y$$

where $T^{-1}Z^{'}X \overset{p}{\rightarrow} D$, $T^{-1}Z'\Omega Z \overset{p}{\rightarrow} \Sigma$ is the asymptotic covariance matrix of the (sample) moment conditions and $V(\hat{\beta}) = (D^{'}\Sigma^{-1}D)^{-1}$ is the asymptotic covariance matrix of the estimator. The matrices D and Σ can be partitioned as follows

$$D = \begin{bmatrix} D_{XX} \\ D_{XZ} \end{bmatrix} \quad \text{and} \quad \Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XZ} \\ \Sigma_{ZX} & \Sigma_{ZZ} \end{bmatrix}.$$

The sub-matrix D_{XZ} contains elements associated with the additional instrumental variables. Based on this representation the estimator that does not make use of the additional instrumental variables in Z_z and coincides with the generalised least squares (GLS) estimator $\hat{\beta}_{GLS} = (X^{'}\Omega^{-1}X)^{-1}X^{'}\Omega^{-1}y$ has the following form

$$\tilde{\beta} = (X' Z_X \Sigma_{XX}^{-1} Z_X' X)^{-1} X' Z_X \Sigma_{XX}^{-1} Z_X' y .$$

It can be shown that the use of additional instrumental variables leads to efficiency gains such that $V(\tilde{\beta}) \geq V(\hat{\beta})$, if and only if (see, Peracchi, 2000, Theorem 11.6)

$$D_{XZ} \geq \Sigma_{ZX} \Sigma_{XX}^{-1} D_{XX} . \tag{2}$$

2. Mean squared prediction error

Next, consider the problem of generating forecasts for h future periods at time T in the context of model (1)

$$y^* = X^* \beta + \varepsilon^*$$

where $y^* = (y_{T+h}, y_{T+h-1}, ..., y_{T+1})'$ is the vector of future values of the dependent variable, X^* is a $h \times K_1$ matrix with the future values of the explanatory variables and ε^* is a $h \times 1$ vector of future disturbances. The variance-covariance matrix of the past and future disturbances can be represented as follows (see Judge et. al. 1985, p. 315)

$$E\left[\begin{pmatrix} \varepsilon \\ \varepsilon^* \end{pmatrix} \begin{pmatrix} \varepsilon & \varepsilon^* \end{pmatrix}\right] = \sigma^2 \begin{bmatrix} \Omega & V \\ V' & \Omega^* \end{bmatrix} = \sigma^2 G . \tag{3}$$

When β , G and the future values of the explanatory variables in X^* are known, Goldberger (1962) proved that the predictor with the minimum mean squared prediction error (MSPE) has the following form

$$y^* = X^*\beta + V'\Omega^{-1}(y - X\beta) \tag{4}$$

with MSPE

$$MSPE(y^*) = E[(y^* - y^*)(y^* - y^*)'] = \Omega^* - V'\Omega^{-1}V\ .$$

When the parameter vector β has to be estimated the predictor in (4) becomes

$$\hat{y}^* = X^*\hat{\beta} + V'\Omega^{-1}(y - X\hat{\beta})\,.$$

When the ordinary least squares (OLS) estimator is used the MSPE is equal to

$$MSPE(\hat{y}^*_{OLS}) = \Omega^* + X^*(X'X)^{-1}X'X(X'X)^{-1} - X^*(X'X)^{-1}X'V - V'X(X'X)^{-1}X^{*'}.$$

The GLS estimator gives the best linear unbiased predictor of y^* with the following MSPE (see Judge et al. 1985, p. 316)

$$\begin{aligned} MSPE(\hat{y}^*_{GLS}) = \Omega^* + X^*(X'\Omega^{-1}X^*)^{-1}X^{*'} - V'(\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X^*)^{-1}X'\Omega^{-1})V \\ - X^*(X'\Omega^{-1}X^*)^{-1}X'\Omega^{-1}V - V'\Omega^{-1}X(X'\Omega^{-1}X^*)^{-1}X^{*'}. \end{aligned}$$

Fang and Koreisha (2004) proposed the following general class of predictors, which includes the OLS and GLS predictors as special cases

$$\hat{y}^* = X^*\hat{\beta} + C'(y - X\hat{\beta})$$

where $\hat{\beta} = (X'\Xi^{-1}X)^{-1}X'\Xi^{-1}y$ and C' is an $(h \times T)$ matrix. The corresponding general form of the mean squared prediction error is

$$MSPE(\hat{y}^*) = \Omega^* - D\Omega^{-1}D' - DV - V'D'$$

where $D = X^*A + C'(I - XA)$ and $A = (X'\Xi^{-1}X)^{-1}X'\Xi^{-1}$. The OLS predictor can be obtained by setting $C = 0$ and $\Xi = I$, while the GLS predictor can be obtained by setting $\Xi = \Omega$ and $C' = V'\Omega^{-1}$.

3. Reducing the mean squared prediction error with instrumental variables

In this section, the mean squared prediction error of the linear regression model is examined when instrumental variables are used in the estimation procedure. It is shown that increasing the number of instruments can reduce the MSPE subject to the validity of the efficiency condition in (2).

Proposition: Let y, $Z = [Z_X, Z_Z]$ and ε be stationary stochastic processes and consider the following estimators of the parameter vector β in model (1) subject to the standard orthogonality and identifiability conditions

$$\hat{\beta} = (X'Z\Sigma^{-1}Z'X)^{-1}X'Z\Sigma^{-1}Z'y \quad \text{and} \quad \tilde{\beta} = (X'Z_X\Sigma_{XX}^{-1}Z_X'X)^{-1}X'Z_X\Sigma_{XX}^{-1}Z_X'y.$$

Assuming that condition (2) holds and the variance-covariance matrix of the past and future disturbances in (3) is known, the following inequality is true

$$MSPE(\hat{y}^*, \tilde{\beta}) \geq MSPE(\hat{y}^*, \hat{\beta}).$$

Proof: Fang and Koreisha (2004), prove that

$$y - X\hat{\beta} = (I - XA)\varepsilon \quad \text{and} \quad \hat{\beta} - \beta = A\varepsilon .$$

If the estimator $\hat{\beta}$ is used A is equal to $(X^{'} Z \Sigma^{-1} Z^{'} X)^{-1} X^{'} Z \Sigma^{-1} Z^{'}$; similarly if the estimator $\tilde{\beta}$ is used A is equal to $(X^{'} Z_X \Sigma_{XX}^{-1} Z_X^{'} X)^{-1} X^{'} Z_X \Sigma_{XX}^{-1} Z_X^{'}$. When using the estimator $\hat{\beta}$ the following expression can be derived for the $MSPE\,(\hat{y}, \hat{\beta}) = E[(\hat{y}^* - y^*)(\hat{y}^* - y^*)^{'}]$ of the predictor $\hat{y}^* = X^* \hat{\beta} + C'(y - X\hat{\beta})$ after substituting equality $y - X\hat{\beta} = (I - XA)\varepsilon$,

$$E\{X^* \hat{\beta} + C^{'}(I - XA)\varepsilon - (X^* \beta + \varepsilon^*)]\,[X^* \hat{\beta} + C^{'}(I - XA)\varepsilon - (X^* \beta + \varepsilon^*)]^{'}\} .$$

By rearranging we get

$$E[(X^* A\varepsilon + C^{'} I\varepsilon - C^{'} XA\varepsilon - \varepsilon^*)\,(X^* A\varepsilon + C^{'} I\varepsilon - C^{'} XA\varepsilon - \varepsilon^*)^{'}] .$$

Substituting equality $\hat{\beta} - \beta = A\varepsilon$ we get the following form

$$E\{[(X^* - C^{'} X)(\hat{\beta} - \beta) + C^{'} I\varepsilon - \varepsilon^*]\,[(X^* - C^{'} X)(\hat{\beta} - \beta) + C^{'} I\varepsilon - \varepsilon^*]^{'}\}$$

which can also be written as follows

$$(X^* - C^{'} X)\,E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)^{'}](X^* - C^{'} X)^{'} + C^{'} I\,E(\varepsilon\varepsilon^{'})\,IC - C^{'} I\,E(\varepsilon\varepsilon^{*'})$$
$$- E(\varepsilon^* \varepsilon^{'})IC + E(\varepsilon^* \varepsilon^{*'}).$$

The MSPE depends on the variance of the estimator $E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)]$ and the elements of the variance-covariance matrix of the past and future disturbances. It then follows that since by assumption inequality (2) holds, it is also true that $E(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)^{'} \geq E(\hat{\beta}-\beta)(\hat{\beta}-\beta)^{'}$ and therefore

$$MSPE(\hat{y}^{*},\tilde{\beta}) \geq MSPE(\hat{y}^{*},\hat{\beta}) \ \square$$

References

Cragg, J. (1983) "More efficient estimation in the presence of heteroskedasticity of unknown form," *Econometrica*, 51: 751–763.

Fang, Y. and Koreisha, S.G. (2004) "Forecasting with serially correlated regression models," *Journal of Statistical Computation and Simulation*, 74: 625-649.

Goldberger, A.S. (1962) "Best linear unbiased prediction in the generalized linear regression model," *Journal of the American Statistical Association*, 57: 369-375.

Johnston, J. and DiNardo, J. (1997) *Econometric Methods* (4th ed.), Singapore: McGraw-Hill.

Judge, G.G., Giffiths, W.E., Hill, C.R., Lutkepohl, H. and Lee, T.C. (1985) *The Theory and Practice of Econometrics* (2nd ed.), New York: John Wiley & Sons Inc.

Peracchi, F. (2000)*Econometrics*, Chichester: John Wiley & Sons.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data* (2nd ed.), Cambridge, MA: MIT press.