

# ADDING REGRESSORS TO OBTAIN EFFICIENCY\*

Sung Jae Jun<sup>†</sup> and Joris Pinkse<sup>‡</sup>  
The Pennsylvania State University

April 2008

## Abstract

It is well-known that in standard linear regression models with i.i.d. data and homoskedasticity, adding ‘irrelevant regressors’ hurts (asymptotic) efficiency unless such irrelevant regressors are orthogonal to the remaining ones. But under (conditional) heteroskedasticity ‘irrelevant regressors’ can always be found such that the ordinary least squares estimator in the long model is as efficient as the generalized least squares estimator in the short model!

---

\*We thank the coeditor and two anonymous referees for their helpful comments. We thank the Human Capital Foundation for their support. Joris Pinkse is an extramural fellow of the Center for Economic Research at Tilburg University

<sup>†</sup>(corresponding author) Department of Economics, The Pennsylvania State University, 608 Kern Graduate Building, University Park PA 16802, sjun@psu.edu

<sup>‡</sup>joris@psu.edu

# 1 Introduction

In standard introductory econometrics texts (see e.g. section 8.4.3 of Greene, 2000) it is explained that with i.i.d. data and under the assumption of (conditional) homoskedasticity adding ‘irrelevant’ regressors (i.e. regressors whose regression coefficients are zero) reduces (asymptotic) efficiency unless such additional regressors are orthogonal to those already present. It is not entirely surprising that with (conditional) heteroskedasticity examples can be found such that the efficiency of the ordinary least squares (OLS) estimator increases if irrelevant regressors are added. What we do find surprising is that irrelevant regressors can always be found such that the OLS estimator in the ‘long’ model is as efficient as the GLS estimator in the ‘short’ model.<sup>1</sup> Such additional regressors depend on unknown population quantities. We show that any generalized least squares (GLS) estimator, be it parametric or semiparametric (e.g. Delgado (1992) and Robinson (1987)) can be interpreted as an OLS estimator in a model with additional regressors.

## 2 Main Results

Consider the linear regression model

$$y_i = x_i' \beta_0 + u_i, \quad i = 1, \dots, n, \quad (1)$$

where  $\{(y_i, x_i)\}$  is i.i.d.,  $E[u_1|x_1] = 0$  a.s.,  $0 < E[x_1 x_1'] < \infty$ , and for some  $0 < c < 1$ ,  $c \leq V(u_1|x_1) \leq 1/c$  a.s.. It is well-known that the GLS estimator  $\hat{\beta}_G$  in (1) has asymptotic variance equal to  $\mathcal{V}_G = (E[x_1 x_1' / \sigma_1^2])^{-1}$ , where  $\sigma_i^2 = V[u_i|x_i]$ .

Now consider the ‘long’ model

$$y_i = x_i' \beta_0 + z_i' \theta_0 + u_i, \quad i = 1, \dots, n, \quad (2)$$

where  $z_i = z(x_i)$  for some function  $z$ .<sup>2</sup> Suppose we estimate (2) using OLS yielding an estimator  $\hat{\beta}_L$  of  $\beta_0$  with asymptotic variance  $\mathcal{V}_L$ . It is well-known that  $\hat{\beta}_L$  can be no more efficient than the

---

<sup>1</sup>For interesting mean square error comparisons between long and short regressions, see Magnus and Durbin (1999).

<sup>2</sup>Note that  $\theta_0 = 0$ , knowledge which is ignored in (2).

GLS estimator of  $\beta_0$  in (2), which in turn is no more efficient than  $\hat{\beta}_G$ . Hence  $\mathcal{V}_G = \mathcal{V}_L$  implies that all three of these estimators are equally efficient.

**Theorem 1** *If the OLS estimator of  $\beta_0$  in (1) has variance  $\mathcal{V}_O \neq \mathcal{V}_G$  then the function  $z$  can be chosen such that  $\mathcal{V}_L = \mathcal{V}_G$ .*

To get some intuition for the result stated in theorem 1, consider the case in which  $z_i$  can be taken to have the same dimension as  $x_i$ . Then

$$y_i = (x_i - z_i)' \beta_0 + z_i' (\beta_0 + \theta_0) + u_i = b_i' \beta_0 + z_i' \delta_0 + u_i, \quad i = 1, \dots, n, \quad (3)$$

for  $b_i = x_i - z_i$  and  $\delta_0 = \beta_0 + \theta_0$ . If  $z_i$  is chosen such that  $E[(x_i - z_i)z_i'] = E[b_i(x_i - b_i)'] = 0$ , then by the Frisch–Waugh–Lovell theorem (Davidson and MacKinnon 2004, section 2.4)  $\mathcal{V}_L$  is the asymptotic variance of the OLS estimator of  $\beta_0$  in a regression of  $y_i$  on  $\tilde{x}_i$  and  $z_i$ , where  $\tilde{x}_i = x_i - E[x_i z_i'] (E[z_i z_i'])^{-1} z_i = x_i - z_i = b_i$ , i.e.  $\mathcal{V}_L$  is the asymptotic variance of the OLS estimator of  $\beta_0$  in (3). Hence

$$\mathcal{V}_L = (E[b_i b_i'])^{-1} E[\sigma_i^2 b_i b_i'] (E[b_i b_i'])^{-1} = (E[b_i x_i'])^{-1} E[\sigma_i^2 b_i b_i'] (E[x_i b_i'])^{-1}. \quad (4)$$

The right-most variance matrix in (4) equals  $\mathcal{V}_G$  whenever  $b_i = C x_i / \sigma_i^2$  for any nonrandom invertible matrix  $C$ .<sup>3</sup> Note however that  $C = E[x_i x_i' / \sigma_i^2] (E[x_i x_i' / \sigma_i^4])^{-1}$  is the only choice that achieves  $E[b_i(x_i - b_i)'] = 0$ . So doing OLS in (2) with  $z_i = x_i - C x_i / \sigma_i^2$  results in  $\mathcal{V}_L = \mathcal{V}_G$ ; the only concern is that this choice of  $z_i$  can belong to a lower dimensional subspace a.s. (the  $z_i$  vector could be collinear), an issue which is addressed in the proof.

Now suppose that one wishes to implement this procedure. One estimates the conditional variances  $\sigma_i^2$ , puts the estimates in a diagonal matrix  $\hat{\Sigma}$  and creates the matrix  $\hat{Z} = [\hat{z}_1, \dots, \hat{z}_n]'$  where  $\hat{z}_i = \hat{z}(x_i)$  and  $\hat{z}$  is like  $z$  but with all population quantities replaced with their sample analogs. Let  $\hat{\beta}_{FG}$  denote the feasible GLS estimator in (1) and  $\hat{\beta}_{FL}$  denote the estimator of  $\beta_0$  in (2) if  $z_i$  is replaced with  $\hat{z}_i$ . Let further  $\hat{\mathcal{V}}_O = (X'X)^{-1} (X' \hat{\Sigma} X) (X'X)^{-1}$  and  $\hat{\mathcal{V}}_G = (X' \hat{\Sigma}^{-1} X)^{-1}$ .

<sup>3</sup>There are in fact no choices for  $b_i$  which differ from  $C x_i / \sigma_i^2$  with positive probability.

**Theorem 2** *If  $\hat{\mathcal{V}}_O \neq \hat{\mathcal{V}}_G$  then  $\hat{\beta}_{FG} = \hat{\beta}_{FL}$ .*

Please note that theorem 2 does not require that the condition of theorem 1 is satisfied. In particular, even if  $\sigma_i^2$  is constant a.s., the condition of theorem 2 may still be satisfied. Since the probability that  $\hat{\mathcal{V}}_O = \hat{\mathcal{V}}_G$  is zero with most weighted least squares procedures, there is no need to test for homoskedasticity ex ante.

In view of theorem 2, the practical usefulness of theorem 1 is limited. However, we believe that most econometricians will find theorem 1 counterintuitive and the proofs of both theorems instructive. Moreover, it follows from the proof of theorem 2 that the class of estimators that involve adding regressors is larger than the class of weighted least squares estimators. So it is plausible that estimators can be constructed that have the same limiting distribution as optimal weighted least squares estimators and work better in practice. It is moreover likely that adding regressors can be used to improve efficiency (possibly including second order efficiency) in more general moment conditions models.

## References Cited:

Davidson, Russell and James G. MacKinnon (2004) *Econometric Theory and Methods*. New York: Oxford University Press.

Delgado, Miguel (1992) Semiparametric generalized least squares in the multivariate nonlinear regression model. *Econometric Theory* 8, 203–240.

Greene, William H. (2000) *Econometric Analysis*, fourth edition. Upper Saddle River NJ: Prentice Hall.

Magnus, Jan R. and James Durbin (1999) “Estimation of regression coefficients of interest when other regression coefficients are of no interest,” *Econometrica* 67, 639–643.

Robinson, Peter M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875–891.

Strang, Gilbert (1988) *Linear algebra and its applications*, Belmont CA: Thomson.

**Proof of theorem 1:**

Drop the  $i$ -subscript throughout. Let  $b = Cx/\sigma^2$  with  $C = E[xx'/\sigma^2](E[xx'/\sigma^4])^{-1}$  and write  $T = E[(x-b)(x-b)'] = ADA'$ , where  $D$  is a diagonal matrix containing the positive eigenvalues of  $T$  and  $A$  is an orthogonal matrix containing eigenvectors corresponding to  $D$ .<sup>4</sup>

Then the solution is  $z = A'(x-b)$ . To see this, note that  $E[b(x-b)'] = 0$ ,  $E[zz'] = D > 0$ , and define

$$\begin{aligned}\tilde{x} &= x - E[xz'](E[zz'])^{-1}z = x - E[x(x-b)']A(A'E[(x-b)(x-b)']A)^{-1}A'(x-b) \\ &= x - TA(A'TA)^{-1}A'(x-b) = b + (I - AA')(x-b) = b \text{ a.s.,}\end{aligned}$$

since  $E[(I-AA')(x-b)(x-b)'(I-AA)'] = 0$ . Then by the Frisch–Waugh–Lovell theorem (Davidson and MacKinnon 2004, section 2.4), we have

$$\begin{aligned}\mathcal{V}_L &= (E[\tilde{x}\tilde{x}'])^{-1}E[\sigma^2\tilde{x}\tilde{x}'](E[\tilde{x}\tilde{x}'])^{-1} = (C')^{-1}(E[xx'/\sigma^4])^{-1}E[xx'/\sigma^2](E[xx'/\sigma^4])^{-1}C^{-1} \\ &= (E[xx'/\sigma^2])^{-1}. \quad \square\end{aligned}$$

**Proof of theorem 2:**

Let  $\hat{B} = \hat{\Sigma}^{-1}X\hat{C}'$  with  $\hat{C} = (X'\hat{\Sigma}^{-1}X)(X'\hat{\Sigma}^{-2}X)^{-1}$ , let  $\hat{A}$  have orthonormal columns which contain the eigenvectors corresponding to the positive eigenvalues of  $(X-\hat{B})'(X-\hat{B})$  and let  $\hat{Z} = (X-\hat{B})\hat{A}$ .<sup>5</sup> Noting that  $(X-\hat{B})'\hat{B} = 0$ , if  $M_{\hat{Z}} = I - P_{\hat{Z}} = I - \hat{Z}(\hat{Z}'\hat{Z})^{-1}\hat{Z}'$  then

$$\begin{aligned}P_{\hat{Z}}X &= (X-\hat{B})\hat{A}(\hat{A}'(X-\hat{B})'(X-\hat{B})\hat{A})^{-1}\hat{A}'(X-\hat{B})'X \\ &= (X-\hat{B})\hat{A}(\hat{A}'(X-\hat{B})'(X-\hat{B})\hat{A})^{-1}\hat{A}'(X-\hat{B})'(X-\hat{B}) = (X-\hat{B})\hat{A}\hat{A}' = X-\hat{B} \text{ a.s.,}\end{aligned}$$

by the definition of  $\hat{A}$ . But then

$$\begin{aligned}\hat{\beta}_{FL} &= (X'M_{\hat{Z}}X)^{-1}X'M_{\hat{Z}}y = (\hat{B}'\hat{B})^{-1}\hat{B}'y = (\hat{C}')^{-1}(X'\hat{\Sigma}^{-2}X)^{-1}X'\hat{\Sigma}^{-1}y \\ &= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y. \quad \square\end{aligned}$$

---

<sup>4</sup> $T$  must have at least one positive eigenvalue, because otherwise  $x = Cx/\sigma^2$  a.s. and  $\mathcal{V}_O = (E[xx'])^{-1}E[\sigma^2xx'](E[xx'])^{-1} = (CE[xx'/\sigma^2])^{-1}CE[xx'/\sigma^2]C'(E[xx'/\sigma^2]C')^{-1} = \mathcal{V}_G$ .

<sup>5</sup> $\hat{A}$  exists, because if  $X = \hat{B}$ , then  $\hat{\mathcal{V}}_O = (X'X)^{-1}(X'\hat{\Sigma}X)(X'X)^{-1} = (\hat{C}'X'\hat{\Sigma}^{-1}X)^{-1}\hat{C}'X'\hat{\Sigma}^{-1}X\hat{C}(X'\hat{\Sigma}^{-1}X\hat{C})^{-1} = \hat{\mathcal{V}}_G$ .