Econometric Theory, **25**, 2009, 298–301. Printed in the United States of America.
doi:10.1017/S0266466608090567

ADDING REGRESSORS TO OBTAIN EFFICIENCY

SUNG JAE JUN AND JORIS PINKSE
Pennsylvania State University

It is well known that in standard linear regression models with independent and identically distributed data and homoskedasticity, adding "irrelevant regressors" hurts (asymptotic) efficiency unless such irrelevant regressors are orthogonal to the remaining regressors. But we have found that under (conditional) heteroskedasticity "irrelevant regressors" can always be found such that one can achieve the asymptotic variance of the generalized least squares estimator by adding the "irrelevant regressors" to the model.

1. INTRODUCTION

In standard introductory econometrics texts (see, e.g., Greene, 2000, Sect. 8.4.3) it is explained that with independent and identically distributed (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 extended model (i.e., the model including the irrelevant regressors) is as efficient as the generalized least squares (GLS) estimator in the reference model (i.e., the model without the irrelevant regressors).[1] Adding regressors to a model changes the estimates of the coefficients β_0 on the existing regressors. The nature of such changes depends on the relationship between existing and added regressors. If the additional regressors are restricted to be fixed functions of the existing regressors corresponding to the same observation, then such functions can be chosen such that the asymptotic variance of the estimates of β_0 equals that of the GLS estimator. If the (unknown) functions are replaced with their sample analogues (such that the additional regressors depend on all data), then a more general result obtains: any feasible GLS estimator, be it parametric (e.g., Aitken, 1935; Harvey,

We thank the co-editor and three 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. Address correspondence to Sung Jae Jun, CAPCP, Department of Economics, 608 Kern Graduate Bldg., University Park, PA 16802, USA; e-mail: sjun@psu.edu.

 © 2009 Cambridge University Press 0266-4666/09 $15.00

https://doi.org/10.1017/S0266466608090567 Published online by Cambridge University Press

1976) or semiparametric (e.g., Delgado, 1992; Robinson, 1987), can be interpreted as an OLS estimator in a model with additional regressors.

2. MAIN RESULTS

Consider the linear regression model

$$y_i = x_i'\beta_0 + u_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $\{(y_i, x_i)\}$ is i.i.d., $\mathrm{E}[u_1|x_1] = 0$ almost surely (a.s.), $0 < \mathrm{E}[x_1 x_1'] < \infty$, 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 $\mathcal{V}_G = \left(\mathrm{E}[x_1 x_1'/\sigma_1^2]\right)^{-1}$, where $\sigma_i^2 = V[u_i|x_i]$. Now consider the extended model

$$y_i = x_i'\beta_0 + z_i'\theta_0 + u_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $z_i = z(x_i)$ for some function z.[2] Suppose we estimate (2) using OLS yielding an estimator $\hat{\beta}_{\mathrm{E}}$ of β_0 with asymptotic variance $\mathcal{V}_{\mathrm{E}}$. It is well known that $\hat{\beta}_{\mathrm{E}}$ can be no more efficient than the GLS estimator of β_0 in (2), which in turn is no more efficient than $\hat{\beta}_G$. Hence $\mathcal{V}_G = \mathcal{V}_{\mathrm{E}}$ implies that all three of these estimators are equally efficient.

THEOREM 1. *If the OLS estimator of β_0 in (1) has variance $\mathcal{V}_O \neq \mathcal{V}_G$ then the function z can be chosen such that $\mathcal{V}_{\mathrm{E}} = \mathcal{V}_G$.*

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

$$y_i = (x_i - z_i)'\beta_0 + z_i'(\beta_0 + \theta_0) + u = b_i'\beta_0 + z_i'\delta_0 + u_i, \qquad i = 1, \ldots, n, \tag{3}$$

for $b_i = x_i - z_i$ and $\delta_0 = \beta_0 + \theta_0$. If z_i is chosen such that $\mathrm{E}[(x_i - z_i)z_i'] = \mathrm{E}[b_i(x_i - b_i)'] = 0$, then by the Frisch–Waugh–Lovell theorem (Davidson and MacKinnon, 2004, Sect. 2.4) $\mathcal{V}_{\mathrm{E}}$ is the asymptotic variance of the OLS estimator of β_0 in a regression of y_i on $\tilde{x}_i$ and z_i, where $\tilde{x}_i = x_i - \mathrm{E}[x_i z_i'](\mathrm{E}[z_i z_i'])^{-1} z_i = x_i - z_i = b_i$, i.e., $\mathcal{V}_{\mathrm{E}}$ is the asymptotic variance of the OLS estimator of β_0 in (3). Hence

$$\mathcal{V}_{\mathrm{E}} = \left(\mathrm{E}[b_i b_i']\right)^{-1}\mathrm{E}[\sigma_i^2 b_i b_i']\left(\mathrm{E}[b_i b_i']\right)^{-1} = \left(\mathrm{E}[b_i x_i']\right)^{-1}\mathrm{E}[\sigma_i^2 b_i b_i']\left(\mathrm{E}[x_i b_i']\right)^{-1}. \tag{4}$$

The rightmost variance matrix in (4) equals $\mathcal{V}_G$ whenever $b_i = Cx_i/\sigma_i^2$ for any nonrandom invertible matrix C.[3] Note however that $C = \mathrm{E}[x_i x_i'/\sigma_i^2]\left(\mathrm{E}[x_i x_i'/\sigma_i^4]\right)^{-1}$ is the only choice that achieves $\mathrm{E}[b_i(x_i - b_i)'] = 0$. So doing OLS in (2) with $z_i = x_i - Cx_i/\sigma_i^2$ results in $\mathcal{V}_{\mathrm{E}} = \mathcal{V}_G$; 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 that is addressed in the proof in section 3.

https://doi.org/10.1017/S0266466608090567 Published online by Cambridge University Press

Now suppose that one wishes to implement this procedure. One estimates the conditional variances σ_i^2 (e.g., Robinson, 1987), puts the estimates in a diagonal matrix $\hat{\Sigma}$, and creates the matrix $\hat{Z} = [\hat{z}_1, \ldots, \hat{z}_n]'$ where $\hat{z}_i = \hat{z}(x_i)$ and $\hat{z}$ is like z but with all population quantities replaced with their sample analogues. Let $\hat{\beta}_{FG}$ denote the feasible GLS estimator in (1) and $\hat{\beta}_{FE}$ denote the estimator of β_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 $\hat{\mathcal{V}}_G = (X'\hat{\Sigma}^{-1}X)^{-1}$.

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

Theorem 1 allows z_i only to depend on x_i, whereas in Theorem 2 the extra regressor—now an estimated quantity—is a function of all data. As a consequence, in Theorem 1 only the equivalence of the asymptotic variances is obtained, whereas in Theorem 2 the estimators themselves are equal. Note further that Theorem 2 does not require that the condition of Theorem 1 is satisfied. In particular, even if σ_i^2 is constant a.s., the condition of Theorem 2 may still be satisfied. Because 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.

3. PROOFS

Proof of Theorem 1. Drop the i subscript throughout. Let $b = Cx/\sigma^2$ with $C = \mathrm{E}[xx'/\sigma^2](\mathrm{E}[xx'/\sigma^4])^{-1}$ and write $T = \mathrm{E}[(x-b)(x-b)'] = ADA'$, where D is a diagonal matrix containing the positive eigenvalues of T and A is an orthogonal matrix containing eigenvectors corresponding to D.[4] Then the solution is $z = A'(x-b)$. To see this, note that $\mathrm{E}[b(x-b)'] = 0$, $\mathrm{E}[zz'] = D > 0$, and define

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

because $\mathrm{E}[(I-AA')(x-b)(x-b)'(I-AA')] = 0$. Then by the Frisch–Waugh–Lovell theorem (Davidson and MacKinnon, 2004, Sect. 2.4), we have

https://doi.org/10.1017/S0266466608090567 Published online by Cambridge University Press

$$\mathcal{V}_{\mathrm{E}} = \left(\mathrm{E}[\tilde{x}\tilde{x}']\right)^{-1}\mathrm{E}\left[\sigma^2\tilde{x}\tilde{x}'\right]\left(\mathrm{E}[\tilde{x}\tilde{x}']\right)^{-1}$$

$$= (C')^{-1}\left(\mathrm{E}\left[xx'/\sigma^4\right]\right)^{-1}\mathrm{E}\left[xx'/\sigma^2\right]\left(\mathrm{E}\left[xx'/\sigma^4\right]\right)^{-1}C^{-1}$$

$$= \left(\mathrm{E}\left[xx'/\sigma^2\right]\right)^{-1}. \quad \blacksquare$$

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 that contain the eigenvectors corresponding to the positive eigenvalues of $(X-\hat{B})'(X-\hat{B})$, and let $\hat{Z} = (X-\hat{B})\hat{A}$.[5] 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

$$P_{\hat{Z}}X = (X-\hat{B})\hat{A}\left(\hat{A}'(X-\hat{B})'(X-\hat{B})\hat{A}\right)^{-1}\hat{A}'(X-\hat{B})'X$$

$$= (X-\hat{B})\hat{A}\left(\hat{A}'(X-\hat{B})'(X-\hat{B})\hat{A}\right)^{-1}\hat{A}'(X-\hat{B})'(X-\hat{B})$$

$$= (X-\hat{B})\hat{A}\hat{A}' = X-\hat{B} \quad \text{a.s.},$$

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

$$\hat{\beta}_{FE} = (X'M_{\hat{Z}}X)^{-1}X'M_{\hat{Z}}y = (\hat{B}'\hat{B})^{-1}\hat{B}'y = (\hat{C}')^{-1}(X'\hat{\Sigma}^{-2}X)^{-1}X'\hat{\Sigma}^{-1}y$$

$$= (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}y. \quad \blacksquare$$

NOTES

1. For interesting mean square error comparisons between these two models, see Magnus and Durbin (1999).

2. Note that $\theta_0 = 0$, knowledge that is ignored in (2).

3. There are in fact no choices for b_i that differ from Cx_i/σ_i^2 with positive probability.

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

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

REFERENCES

Aitken, A. (1935) On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh* 55, 42–48.

Davidson, R. & J.G. MacKinnon (2004) *Econometric Theory and Methods*. Oxford University Press.

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

Greene, W.H. (2000) *Econometric Analysis*, 4th ed. Prentice-Hall.

Harvey, A. (1976) Estimating regression models with multiplicative heteroscedasticity. *Econometrica* 44, 461–465.

Magnus, J.R. & J. Durbin (1999) Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica* 67, 639–643.

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

https://doi.org/10.1017/S0266466608090567 Published online by Cambridge University Press