Linear models with nonscalar covariance matrix and generalized least squares [*]

Jean-Marie Dufour [†]
McGill University

First version: March 2002
Revised: August 2011
This version: August 2011
Compiled: August 20, 2011, 12:06

[*] This work was supported by the William Dow Chair in Political Economy (McGill University), the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Bank of Canada (Research Fellowship), a Guggenheim Fellowship, a Konrad-Adenauer Fellowship (Alexander-von-Humboldt Foundation, Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, and the Fonds de recherche sur la société et la culture (Québec).

[†] William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en économie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 519, 855 Sherbrooke Street West, Montréal, Québec H3A 2T7, Canada. TEL: (1) 514 398 8879; FAX: (1) 514 398 4938; e-mail: jean-marie.dufour@mcgill.ca . Web page: http://www.jeanmariedufour.com

Contents

1. Generalized least squares **1**

1.1. Best linear unbiased estimator 1

1.2. Gaussian case 4

2. Estimation with heteroskedasticity **5**

2.1. Known variance structure 5

2.2. Unknown variance structure 5

1. Generalized least squares

1.1. Best linear unbiased estimator

$$y = X\beta + u \tag{1.1}$$

where y is a $T \times 1$ vector of observations on a dependent variable, X is a $T \times k$ nonstochastic matrix of rank k, and u is a $T \times 1$ vector of disturbances (errors) such that

$$\begin{aligned} \mathsf{E}(u) &= 0 \\ \mathsf{V}(u) &= \sigma^2 V \end{aligned} \tag{1.2}$$

and V is a known $T \times T$ positive definite matrix. Then the least-squares estimator

$$\hat{\beta} = (X'X)^{-1}X'y \tag{1.3}$$

is unbiased but does not have minimal variance. The covariance matrix of $\hat{\beta}$ is

$$\mathsf{V}(\hat{\beta}) = \sigma^2(X'X)^{-1}X'VX(X'X)^{-1} \tag{1.4}$$

so that the usual formula

$$\mathsf{V}(\hat{\beta}) = \sigma^2(X'X)^{-1} \tag{1.5}$$

is not valid.

The fact V is positive definite entails that $|V| \neq 0$, so there is no perfect correlation between the disturbances. Further, there exists a nonsingular $T \times T$ matrix P such that

$$PVP' = I_T\,, \tag{1.6}$$

$$(P')^{-1}V^{-1}P^{-1} = (PVP')^{-1} = I_T\,. \tag{1.7}$$

Multiply both sides of (1.1) by P :

$$Py = PX\beta + Pu\,. \tag{1.8}$$

We get in this way the transformed model

$$y_* = X_*\beta + u_* \tag{1.9}$$

where

$$y_* = Py, \quad X_* = PX, \quad u_* = Pu \tag{1.10}$$

$$\begin{aligned} \mathsf{E}(u_*) &= 0, & (1.11)\\ \mathsf{V}(u_*) &= \mathsf{E}\left[Puu'P'\right] = \sigma^2 PVP' = \sigma^2 I_T. & (1.12)\end{aligned}$$

Then

$$\hat{\beta}_G = \left(X_*'X_*\right)^{-1} X_*'y_* \tag{1.13}$$

is the best linear unbiased estimator of β :

$$\begin{aligned} \mathsf{E}(\hat{\beta}_G) &= \beta \\ \mathsf{V}(\hat{\beta}_G) &= \sigma^2(X_*'X_*)^{-1}. \end{aligned} \tag{1.14}$$

We can also write:

$$\hat{\beta}_G = (X'P'PX)^{-1}X'P'Py = \left(X'V^{-1}X\right)^{-1}X'V^{-1}y \tag{1.15}$$

for

$$\begin{aligned} PVP' = I_T \ &\Rightarrow V = P^{-1}(P')^{-1} = (P'P)^{-1} \\ &\Rightarrow V^{-1} = P'P. \end{aligned} \tag{1.16}$$

$\hat{\beta}_G$ is called the generalized least squares estimator of β :

$$\begin{aligned} \mathsf{E}\left(\hat{\beta}_G\right) &= \beta, \\ \mathsf{V}\left(\hat{\beta}_G\right) &= \sigma^2\left(X_*'X_*\right) = \sigma^2\left(X'V^{-1}X\right)^{-1}. \end{aligned} \tag{1.17}$$

We know that $\hat{\beta}$ minimizes

$$(y - X\beta)'(y - X\beta). \tag{1.18}$$

Similarly, $\hat{\beta}_G$ minimizes

$$\begin{aligned}
(y_* - X_*\beta)'(y_* - X_*\beta) &= (Py - PX\beta)'(Py - PX\beta) \\
&= (y - X\beta)'P'P(y - X\beta) \\
&= (y - X\beta)'V^{-1}(y - X\beta)
\end{aligned}$$

This is why $\hat{\beta}_G$ is also called a *weighted least squares* estimator of β.

1.2. Gaussian case

Suppose

$$u \sim N\left[0, \sigma^2 V\right] \tag{1.19}$$

Then

$$\hat{\beta}_G \sim N\left[\beta, \sigma^2\left(X'V^{-1}X\right)^{-1}\right] \tag{1.20}$$

is the best mean squares unbiased estimator of β.

We can build tests and confidence intervals in the usual manner by using the transformed model

$$(Py) = (PX)\beta + (Pu) \tag{1.21}$$

instead of

$$y = X\beta + u. \tag{1.22}$$

2. Estimation with heteroskedasticity

2.1. Known variance structure

Suppose

$$\mathsf{E}\left[uu'\right] = \sigma^2 \begin{bmatrix} d_1^2 & 0 & \cdots & 0 \\ 0 & d_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_T^2 \end{bmatrix} = \sigma^2 V\,. \tag{2.1}$$

The variance of each element of u is then

$$\mathsf{V}(u_t) = \sigma_t^2 = d_t^2\sigma^2 \tag{2.2}$$

and we have:

$$y_t = x_t'\beta + u_t\ , \quad t = 1, \ldots, T$$

$$\frac{y_t}{d_t} = \frac{1}{d_t}x_t'\beta + \frac{u_t}{d_t}\ , \quad t = 1, \ldots, T \tag{2.3}$$

$$y_{*t} = x_{*t}'\beta + u_{*t}\ ,\ t = 1, \ldots, T \tag{2.4}$$

$$\mathsf{V}(u_{*t}^2) = \mathsf{V}\left(\frac{u_t}{d_t}\right) = \sigma^2\frac{d_t^2}{d_t^2} = \sigma^2 \tag{2.5}$$

$$P = \begin{bmatrix} 1/d_1 & 0 & \cdots & 0 \\ 0 & 1/d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/d_T \end{bmatrix} \tag{2.6}$$

2.2. Unknown variance structure

It is rare that $d_1, \ldots, d_T$ are known.

It is impossible to estimate $T + k$ parameters with T observations (incodental parameter problem)..

One must make hypotheses on the form of the variance structure.

1. $d_t^2 = c\left(x_{tk}\right)^2$

where x_k is one of the explanatory variables or another variable. Then

$$\begin{aligned}
\frac{y_t}{x_{tk}} &= \frac{1}{x_{tk}} x_t' \beta + \frac{u_t}{x_{tk}}\,, \quad t = 1, \ldots, T \\
\mathsf{V}\left(\frac{u_t}{x_{tk}}\right) &= \sigma^2 c = c\sigma^2
\end{aligned} \tag{2.7}$$

2. $\sigma_t^2 = c\,(\mathsf{E} y_t)^2 = c\,(x_t'\beta)^2$
   Then

$$\frac{y_t}{\mathsf{E}(y_t)} = \frac{1}{\mathsf{E}(y_t)} x_t'\beta + \frac{u_t}{\mathsf{E}(y_t)}\,, \quad t = 1, \ldots, T \tag{2.8}$$

A difficulty here is that $\mathsf{E}(y_t) = x_t'\beta$ is unknown. This suggests a two-step procedure.

1. Estimate β par OLS. This is reasonable because $\hat{\beta}$ is unbiased.
2. The model is then transformed according to:

$$\frac{y_t}{x_t'\hat{\beta}} = \left(\frac{1}{x_t'\hat{\beta}} x_t'\right)\beta + \frac{u_t}{x_t'\hat{\beta}}\,. \tag{2.9}$$

   In this way, the model becomes “approximately homoskedastic”. For T large this leads to efficient estimators and valid tests and confidence intervals.