Quaderni di Statistica
Vol. 7, 2005

Is aggregation ever necessary? *

Luciano Pieraccini
Dipartimento di Economia, Università degli Studi Roma Tre
E-mail: pieracci@uniroma3.it

Summary: The relations among parameters in aggregate and elementary models are investigated and BLU estimators of elementary parameters are used to derive BLU estimators of the aggregate one showing that is possible to switch from former to latter without introducing aggregation bias. On this bases conditions for perfect aggregation are established and a proper measure of goodness of fit is derived both for aggregate and elementary models. It is then shown that, even in the case of perfect aggregation, elementary models have to be preferred to the aggregate one on the ground of the defined goodness of fit criterion.

Keywords: Aggregation, Aggregation bias, Perfect aggregation, BLU Estimators, Goodness of fit.

1. Introduction

The problem of aggregation of economic relations has a very long tradition in Econometrics: starting from the pioneer work of H. Theil (1954) and passing through the contribution of A. Zellner (1962) up to the present day, the topic of aggregation and the consequential topic of aggregation bias has often received the attention of econometricians. In recent times the unification of Europe has given a new impulse to the subject, since the micro-relations for individual country members need to be compared with the macro-relation for the entire UE. It is in this particular contest that the present work has been developed.

---

* This paper has been presented at the 1st Italian Congress on Econometrics and Empirical Economics, Venice, January 24th-25th, 2005.

In a very well known paper Y. Grunfeld and Z. Griliches (1962) argued "Is aggregation necessarily bad?" concluding that under some circumstances aggregation could give rise to a more efficient estimation than disaggregated relations. Our position is completely opposite to that: we are arguing whether the estimation of aggregate models is ever necessary. In point of fact we think that the problem of aggregation is a false problem since we will show that it is possible to switch from the disaggregate models to the aggregate one without introducing aggregation bias.

This result comes out very easily from the investigation of relations among parameters in aggregate and disaggregate models (Section 3), so that it is possible to derive BLU estimators of aggregate parameters on the basis of elementary ones (Section 4). Conditions for perfect aggregation are derived (Section 5) pointing out that, even in more recent times, there have been some misunderstandings about testing for perfect aggregation. Finally (Section 6) a proper measure of goodness of fit is derived both for aggregate and disaggregate models; it is furthermore shown that, even in the case of perfect aggregation, elementary models have to be preferred on the ground of the defined goodness of fit criterion.

2. Establishing notation

Let's consider the following system of linear models

$$\begin{aligned} Y_1 &= X_1\beta_1 + \varepsilon_1 \\ &\vdots \\ Y_c &= X_c\beta_c + \varepsilon_c \\ &\vdots \\ Y_C &= X_C\beta_C + \varepsilon_C \end{aligned} \tag{1}$$

where the endogenous variable Y observed at time t $(t = 1,\ldots,T)$ in country c $(c = 1,\ldots,C)$ is expressed as a linear function of V

exogenous variables X_v $(v = 1,\ldots,V)$, according to the following definitions of vectors and matrices

$$Y_c = (y_{c1},\ldots,y_{ct},\ldots y_{cT})'$$

$T \times 1$ vector of the endogenous variable in country c,

$$X_c = \begin{bmatrix} x_{c11} & \cdots & x_{cV1} \\ \vdots & \ddots & \vdots \\ x_{c1T} & \ldots & x_{cVT} \end{bmatrix}$$

$T \times V$ matrix of exogenous variables in country c,

$$\beta_c = (\beta_{c1},\ldots,\beta_{cv},\ldots,\beta_{cV})'$$

$V \times 1$ vector of coefficients for country c

$$\varepsilon_c = (\varepsilon_{c1},\ldots,\varepsilon_{ct}\ldots,\varepsilon_{cT})'$$

$T \times 1$ vector of disturbances for country c.

Using a standard notation, the system of equations (1) can be written as

$$\tilde{Y} = \tilde{X}\tilde{\beta} + \tilde{\varepsilon} \tag{2}$$

where vectors and matrices are defined as follows

$$\tilde{Y} = (Y_1',\ldots,Y_c',\ldots,Y_C')'$$

$$\tilde{X} = \begin{bmatrix} X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & X_C \end{bmatrix}$$

$$\tilde{\beta} = (\beta_1',\ldots,\beta_c',\ldots,\beta_C')'$$

$$\tilde{\varepsilon} = (\varepsilon_1',\ldots,\varepsilon_c',\ldots,\varepsilon_C')'$$

At time *t,* the system of equations (1) particularises in the following way

$$
\begin{aligned}
y_{1t} &= X_{1t}\beta_1 + \varepsilon_{1t} \\
&\vdots \\
y_{ct} &= X_{ct}\beta_c + \varepsilon_{ct} \\
&\vdots \\
y_{Ct} &= X_{Ct}\beta_C + \varepsilon_{Ct}
\end{aligned}
\tag{3}
$$

where y_{ct} is the *t-th* element of the vector Y_c (i.e. the value assumed by the endogenous variable in country c at time t), X_{ct} is the vector given by the *t-th* row of X_c (i.e. the $1 \times V$ vector of values assumed by the V exogenous variables in country c at time t), β_c is the vector of coefficients for country c and finally ε_{ct} is the disturbance for country c at time t.

The set of C equations (3) at time t can be written in the following way using the same notation as in (2)

$$\tilde{Y}_t = \tilde{X}_t \tilde{\beta} + \tilde{\varepsilon}_t \tag{4}$$

where

$$\tilde{Y}_t = (y_{1t}, \ldots, y_{ct}, \ldots, y_{Ct})'$$

and $\tilde{X}$ is the vector diagonal matrix

$$
\tilde{X}_t = \begin{bmatrix}
X_{1t} & \ldots & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots \\
0 & \ldots & X_{ct} & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & \ldots & X_{Ct}
\end{bmatrix}
$$

The C equations in whatsoever form (1), (2), (3) or (4) will be called *elementary models* or *micro models* to distinguish them from the *aggregate model* or *macro model* that we are now going to introduce.

3. The aggregate model and its relationship to the elementary ones

Defining the following vectors and matrices of sums over countries

$$Y_{\bullet} = (\sum_{c=1}^{C} y_{c1}, \ldots, \sum_{c=1}^{C} y_{ct}, \ldots, \sum_{c=1}^{C} y_{cT})' = (y_{\bullet 1}, \ldots, y_{\bullet t}, \ldots, y_{\bullet T})'$$

$$X_{\bullet} = \begin{bmatrix} \sum_{c=1}^{C} x_{c11} & \cdots & \sum_{c=1}^{C} x_{cV1} \\ \vdots & \ddots & \vdots \\ \sum_{c=!}^{C} x_{c1T} & \cdots & \sum_{c=!}^{C} x_{cVT} \end{bmatrix} = \begin{bmatrix} x_{\bullet 11} & \cdots & x_{\bullet V1} \\ \vdots & \ddots & \vdots \\ x_{\bullet 1T} & \cdots & x_{\bullet VT} \end{bmatrix} = \begin{bmatrix} X_{\bullet 1} \\ \vdots \\ X_{\bullet T} \end{bmatrix}$$

$$\varepsilon_{\bullet} = (\sum_{c=1}^{C} \varepsilon_{c1}, \ldots, \sum_{c=1}^{C} \varepsilon_{ct} \ldots, \sum_{c=1}^{C} \varepsilon_{cT})' = (\varepsilon_{\bullet 1}, \ldots, \varepsilon_{\bullet t} \ldots, \varepsilon_{\bullet T})'$$

and the vectors of the parameters

$$\delta = (\delta_1, \ldots, \delta_v, \ldots \delta_V)'$$

the aggregate model may then be written as

$$Y_{\bullet} = X_{\bullet}\delta + \varepsilon_{\bullet} \tag{5}$$

which at time t becomes

$$y_{\bullet t} = X_{\bullet t}\delta + \varepsilon_{\bullet t} \tag{6}$$

Our aim is to see whether direct estimation of model (5) gives rise to a model which, in some sense, performs better than the estimated elementary ones.

To explore the relations between the parameters β of the elementary models and the parameters δ of the aggregated one, let's consider that $y_{\bullet t}$ can be obtained pre-multiplying the left hand side of equations (4) by the $C \times 1$ row vector of ones $\mathbf{1}' = (1, 1, \ldots, 1)$.

Performing the same operation on the right hand side of equations (4), the following expression is obtained

$$\mathbf{1}'\tilde{Y}_t = \mathbf{1}'\tilde{X}_t\tilde{\beta} + \mathbf{1}'\tilde{\varepsilon}_t \tag{7}$$

that is

$$\sum_{c=1}^{C} y_{ct} = \sum_{c=!}^{C}\sum_{v=1}^{V} x_{cvt}\beta_{cv} + \varepsilon_{\bullet t} = \sum_{v=!}^{V} x_{\bullet vt}\sum_{c=1}^{C}(x_{cvt}/x_{\bullet vt})\beta_{cv} + \varepsilon_{\bullet t}$$

or

$$y_{\bullet t} = \sum_{v=1}^{V} x_{\bullet vt}\sum_{c=1}^{C}\omega_{cvt}\beta_{cv} + \varepsilon_{\bullet t} = \sum_{v=1}^{V} x_{\bullet vt}\gamma_{vt} + \varepsilon_{\bullet t}$$

where ω_{ct} and γ_{vt} are defined as follows

$$\omega_{cvt} = \frac{x_{cvt}}{x_{\bullet vt}} \qquad (\sum_{c=1}^{C}\omega_{cvt} = 1) \tag{8}$$

$$\gamma_{vt} = \sum_{c=1}^{C}\omega_{cvt}\beta_{cv} \tag{9}$$

Setting now

$$\gamma_t = (\gamma_{1t}, \ldots, \gamma_{vt}, \ldots, \gamma_{Vt})'$$

model (7) may be finally written as

$$y_{\bullet t} = X_{\bullet t}\gamma_t + \varepsilon_{\bullet t} \quad . \tag{10}$$

The aggregation of the endogenous variable Y over countries induces an aggregation of the exogenous variables X through parameters γ_t which are weighted averages (9) of the β coefficients of the elementary models. Furthermore, since the weights are time dependent, the vectors γ_t are intrinsically time-varying parameters, depending on the variation of the exogenous variables X over time and over countries in the C elementary models.

4. Estimation

Under the usual condition on disturbances

$$E(\tilde{\varepsilon}) = 0\,; \qquad E(\tilde{\varepsilon}\tilde{\varepsilon}') = \Omega \otimes I$$

with Ω being the $C \times C$ variance covariance matrix of disturbances between elementary models (which is assumed to be constant over time), it is well known that the GLS estimators of parameters of elementary models (Zellner's SUR procedure) are BLU.

Let $\hat{\beta}$ be the $CV \times 1$ GLS vector estimator of elementary models parameter

$$\hat{\beta} = (\tilde{X}'(\Omega^{-1} \otimes I)\tilde{X})^{-1}\tilde{X}'(\Omega^{-1} \otimes I)\tilde{Y} \tag{11}$$

and $\hat{Y}$ the $CT \times 1$ vector of the endogenous variable, the estimated model will then be

$$\hat{Y} = \tilde{X}\hat{\beta}$$

It is also well known that the BLU estimator of a linear combination of parameters is the linear combination of BLU estimators so that

$$\hat{\gamma}_{vt} = \sum_{c=1}^{C} \omega_{cvt}\hat{\beta}_{cv}$$

will be BLU.

It has to be stressed that the estimated aggregate model

$$\hat{y}_{\bullet t} = X_{\bullet t}\hat{\gamma}_t \qquad \forall t \tag{12}$$

will show perfect aggregation since it will be

$$\hat{y}_{\bullet t} = \sum_{c=1}^{C} \hat{y}_{ct} \tag{13}$$

and also

$$e_{\bullet t} = \sum_{c=1}^{C} e_{ct} \tag{14}$$

where as usual e are the residuals from estimated models.

Because of (13) and (14), no aggregation bias will be introduced in switching from the elementary models to the aggregate one.

If, as it usually happens, the Ω matrix is not known, the estimators adopted for both kind of parameters β's and γ's, will be the "feasible" version of GLS.

On the contrary, the direct least squares estimation of aggregate model will generally introduce aggregation bias. This is clearly seen considering that in estimating directly model (5), the model that is really estimated will be of the type

$$Y_{\bullet t} = X_{\bullet t}\delta + X_{\bullet t}(\gamma_t - \delta) + \varepsilon_{\bullet t} \tag{15}$$

or

$$Y_{\bullet t} = X_{\bullet t}\delta + u_t$$

with

$$u_t = X_{\bullet t}(\gamma_t - \delta) + \varepsilon_{\bullet t}$$

This situation will have the following consequences on the new disturbance term u_t

$$\begin{aligned} &E(u_t) = X_{\bullet t}(\gamma_t - \delta) \neq 0 \\ &E(u_t^2) = (\gamma_t - \delta)' X'_{\bullet t} X_{\bullet t}(\gamma_t - \delta) + Var(\varepsilon_{\bullet}) \\ &E(u_t u_\tau) = (\gamma_t - \delta)' X'_{\bullet t} X_{\bullet \tau}(\gamma_\tau - \delta) \neq 0 \qquad \forall t \neq \tau \end{aligned} \tag{16}$$

Because of equations (16) least squares direct estimation of aggregate model will give rise to biased estimators.

From equation (15), it comes out directly that when the following condition holds

$$\gamma_t = \delta \qquad \forall t \tag{17}$$

no aggregation bias will be introduced switching from the elementary models to the aggregate one, and there will be perfect aggregation.

5. Conditions for perfect aggregation

To consider under which hypothesis the direct estimation of the aggregate model will be consistent with the estimation of the elementary ones, it is necessary to investigate the conditions under which (17) holds. We are going to consider three different situations which are the more interesting ones from our point of view.

1 - As Zellner(1962) pointed out, there will be no aggregation bias when the β's are constant over countries, that is when

$$\beta_{vc} = \beta_v \qquad \forall c \tag{18}$$

Under this condition it will in point of fact be

$$\gamma_{vt} = \beta_v \sum_{c=1}^{C} \omega_{cvt} = \beta_v = \delta_v$$

and the weights ω will be irrelevant.

In the same work Zellner proposed a test for perfect aggregation based on this kind of null hypothesis.

2 - In an attempt to postulate less stringent conditions for perfect aggregation, Lee, Pesaran and Pierce(1990) considered the case of the equality of an average of the β's of elementary models to the corresponding parameters of the aggregate one. In our notation it will be

$$\gamma_v = \frac{1}{C} \sum_{c=1}^{C} \beta_{cv} = \delta_v \tag{19}$$

But according to (17), to have perfect aggregation under this condition it has to be

$$\omega_{cvt} = constant \qquad \forall c \quad (20)$$

i.e.

$$\frac{x_{cvt}}{x_{\bullet vt}} = constant \qquad \forall c$$

which is verified if and only if x_{cvt} is constant over countries. It is then evident that this is a very peculiar kind of hypothesis and that it is very much more stringent and less credible then the one proposed by Zellner.

3 - A third condition for perfect aggregation has to be considered, namely when

$$\omega_{cvt} = \omega_{cv} \qquad \forall t \tag{21}$$

i.e. when the weights are constant over times

$$\frac{x_{cvt}}{x_{\bullet vt}} = \omega_{cv} \qquad \forall t$$

This situation will arise when the x_{cvt}'s vary proportionally over countries from time to time, so that an overall increase of $\alpha\%$ in the aggregated variable matches with the same increase in each country. In this case it will be

$$\gamma_v = \sum_{c=1}^{C} \omega_{cv} \beta_v = \delta_v \tag{22}$$

In the same paper, Lee, Pesaran and Pierce (1990) proposed a test for perfect aggregation based on the equality of the parameters of aggregate model to weighted averages of the elementary ones. The test can be properly applied in this context.

6. Measuring goodness of fit

The argument that is sometimes used in favour of direct estimation of the aggregate model is that it will give rise to a better fitting model. Then the point that has to be explored is the way the goodness of fit is measured.

Our strong belief is that any measure of goodness of fit has to be referred to the total variance of the $C \times T$ observations on the endogenous variable Y, that is

$$Var(Y) = \tfrac{1}{CT}\sum_{c=!}^{C}\sum_{t=1}^{T}(y_{ct} - \bar{y})^2$$

where $\bar{y}$ is the overall mean

$$\bar{y} = \tfrac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T} y_{ct}$$

Let now $\bar{y}_t$ be the mean of the observations on the C countries at time *t*

$$\bar{y}_t = \tfrac{1}{C}\sum_{c=1}^{C} y_{ct} = \tfrac{1}{C} y_{\bullet t} \tag{23}$$

and let $\bar{y}_c$ be the mean of the observations over the T times in country *c*

$$\bar{y}_c = \tfrac{1}{T}\sum_{t=1}^{T} y_{ct} \tag{24}$$

With respect to (23) the total variance of Y can be written in the following way

$$\begin{aligned} Var(Y) &= \tfrac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}(y_{ct} - \bar{y}_t)^2 + \tfrac{1}{T}\sum_{t=1}^{T}(\bar{y}_t - \bar{y})^2 \\ &= \sigma_W^2(t) + \sigma_B^2(t) \end{aligned}$$

where $\sigma_W^2(t)$ is the *within times variance* while $\sigma_B^2(t)$ is the *between times variance*.

With the aggregate model we are modelling $\sigma_B^2(t)$ since, dividing both sides of (6) by C, it comes out

$$\tfrac{1}{C} Y_{\bullet t} = \tfrac{1}{C} X_{\bullet t}\delta + \tfrac{1}{C}\varepsilon_{\bullet t}$$

i.e.

$$\bar{y}_t = \bar{X}_t\delta + \bar{\varepsilon}_t$$

where $\overline{X}_t$ is the $1 \times V$ vector of over countries means of exogenous variables, and $\overline{\varepsilon}_t$ is the corresponding mean of disturbances.

On the other hand, the total variance can be partitioned with respect to (24) obtaining

$$\begin{aligned} Var(Y) &= \tfrac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}(y_{ct}-\overline{y}_c)^2 + \tfrac{1}{C}\sum_{c=1}^{C}(\overline{y}_c-\overline{y})^2 \\ &= \sigma_W^2(c) + \sigma_B^2(c) \end{aligned}$$

where $\sigma_W^2(c)$ and $\sigma_B^2(c)$ are the *within countries* and *between countries* variances.

With the elementary models we are then modelling $\sigma_W^2(c)$ since

$$y_{ct} = X_{ct}\beta_c + \varepsilon_{ct}$$

is going to explain the within country variations.

In both cases the appropriate denominator for whatsoever goodness of fit index is the total variance of Y.

Defining

$$\widehat{y}_{\bullet t} = X_{\bullet t}\widehat{\delta}$$

for the aggregate model the measure of goodness of fit is therefore

$$R_{\bullet}^2 = \frac{\sum_{t=1}^{T}(\widehat{y}_{\bullet t} - \overline{y}_{\bullet})^2}{C\sum_{t=1}^{T}\sum_{c=1}^{C}(y_{ct}-\overline{y})^2} \tag{25}$$

While for the elementary models, defining

$$\widehat{y}_{ct} = X_{ct}\widehat{\beta}_c$$

the measure of goodness of fit is

$$R^2 = \frac{\sum_{t=1}^{T}\sum_{c=1}^{C}(\widehat{y}_{ct} - \overline{y}_c)^2}{\sum_{t=1}^{T}\sum_{c=1}^{C}(y_{ct} - \overline{y})^2} \tag{26}$$

The two measures $R_{\bullet}^2$ and R^2 will then be homogeneous and the comparison among them appropriate: that will avoid any mis-understanding on the behaviour of the two models.

It has to be noticed that even in the case of perfect aggregation it will be

$$R^2 \geq R_{\bullet}^2 \tag{27}$$

Taking into account (11), the numerator of (25) becomes

$$\tfrac{1}{C}\sum_{t=1}^{T}(\sum_{c=1}^{C}\widehat{y}_{ct} - \sum_{c=1}^{C}\overline{y}_c)^2 = \tfrac{1}{C}\sum_{t=1}^{T}\left[\sum_{c=1}^{C}(\widehat{y}_{ct} - \overline{y}_c)\right]^2$$

so that it will always be

$$\sum_{t=1}^{T}\sum_{c=1}^{C}(\widehat{y}_{ct} - \overline{y}_c)^2 \geq \tfrac{1}{C}\sum_{t=1}^{T}\left[\sum_{c=1}^{C}(\widehat{y}_{ct} - \overline{y}_c)\right]^2$$

being

$$\sum_{c=1}^{C}(\widehat{y}_{ct} - \overline{y}_c)^2 \geq \tfrac{1}{C}\left[\sum_{c=!}^{C}(\widehat{y}_{ct} - \overline{y}_c)\right]^2 \qquad \forall t$$

The relation (27) shows that, even in the case of perfect aggregation, the elementary models have to be preferred to the aggregate one on the ground of goodness of fit criterion.

7. Conclusions

We have seen that the relation derived among parameters of elementary and aggregate models allows BLU estimation of both of them, switching from former to latter without introducing aggregation bias. The aggregate model parameters are then fully consistent to those of the elementary models, so that predicted values of the former sum up to those of the latter and the same happens for residuals.

On the contrary, direct estimation of aggregate model parameters generates in general aggregation bias. Only in few cases of perfect aggregation direct estimation does not introduce aggregation bias: but we have shown that these are very peculiar cases so that testing for perfect aggregation have to be interpreted in the light of them.

Furthermore, even in presence of perfect aggregation it is shown that the estimation of micro-relations has to be preferred. To this extent, to avoid misunderstanding about the performance of aggregate and elementary models, a proper measure of goodness of fit has been introduced. On its light, it is shown elementary models always to perform better than the aggregate one: in our opinion this result is perfectly in line with the loss of information already pointed out by Orcutt, Watts and Edwards (1968).

"Rebus sic stantibus", we are deeply convinced that there is no reason for estimating an aggregate model instead of elementary ones: this is particularly true in the case of micro-relations relative to country members as counterpart of a macro-relation for all of them. The situation may be different if we allow misspecification in micro-relations as Grunfeld and Griliches (1960) did, but then there is the problem arising from the possibility that the aggregate model could be not properly specified while the elementary ones are exactly specified.

Finally, we want to stress that – in our opinion – the time path of $\hat{\gamma}_t$ can be useful to draw some light on their future evolution and on the future behaviour of the aggregate dependent variable. But this is another story, that should be fully explored.

References

Grunfeld, Y., and Griliches, Z. (1960), Is Aggregation Necessarily Bad?, *Review of Economics and Statistics*, 42, 1-13.

Lee, K. C., Pesaran, M. C., and Pierse, R. G. (1990), Testing for Aggregation Bias in Linear Models, *The Economic Journal (Supplement)*, 100, 137-150.

Orcutt, G. H., Watts, H. W., Edwards, J. B. (1968), Data Aggregation and Information Loss, *The American Economic Review,* 58, 773-787. Theil H. (1954), *Linear Aggregation of Economic Relations*, Amsterdam: North Holland.

Zellner A. (1962), An Efficient method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, *Journal of the American Statistical Association*, 57, 348-368.